INVESTIGATING THE POTENTIAL OF ON-DEMAND RIDE SERVICE AND ITS IMPACT ON MODE CHOICE AND ACCESSIBILITY

A Dissertation
Presented to
The Academic Faculty

by

Fangru Wang

In Partial Fulfillment
of the Requirements for the Degree
Ph.D. in the
City and Regional Planning / College of Design

Georgia Institute of Technology
August of 2018

INVESTIGATING THE POTENTIAL OF ON-DEMAND RIDE SERVICE AND ITS IMPACT ON MODE CHOICE AND ACCESSIBILITY

Approved by:

Dr. Catherine Ross, Advisor
School of City and Regional Planning
Georgia Institute of Technology

Dr. Alex Karner
School of Architecture
The University of Texas at Austin

Dr. Patricia Mokhtarian
School of Civil and Environmental Engineering
Georgia Institute of Technology

Jamie Cochran, FAICP
School of City and Regional Planning
Georgia Institute of Technology

Dr. Bruce Stiftel
School of City and Regional Planning
Georgia Institute of Technology

Date Approved: June 29th, 2018

ACKNOWLEDGEMENTS

Numerous individuals have helped me in developing this dissertation. First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Catherine Ross. Her continuous support and thoughtful feedback have guided me through this academic journey. I also thank her for funding my study and giving me the great opportunity to work on multiple research projects that broaden my horizon as a researcher. I also owe a debt of gratitude to each of my committee members, Dr. Patricia Mokhtarian, Dr. Alex Karner, Dr. Bruce Stiftel, and Ms. Jamie Cochran. My committee members have not only provided me insightful comments that motivate me to widen and deepen my research but also given me endless encouragement and support.

My sincere thanks also go to Dr. Tim Welch who has provided me enlightening feedback and support since the very early stage of developing this dissertation. I thank all the researchers, staff, and fellow students in the Center for Quality Growth and Regional Development (CQGRD), for their constant help and encouragement, especially Peter, Arthi, Angie, Sarah, Amit, David, Abi, Cary, Robby, Laura, Sigal, and Gervais. I also wholeheartedly thank my friends Wenwen, Camille, Haobing, Yongsung, Yunkyung, Ben, and Kevin, for the enlightening collaboration, inspiring discussions, and joyful accompany that make this journey sweet and memorable.

Finally, I could not have made it this far without my family. I am grateful to my dad, Lumin and my mom, Ying for their everlasting love and inspiring thoughts. I thank my husband, Zekai for his constant support, constructive advice, and tireless travel for always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

BG	Block group
DT	Decision tree
FHWA	Federal Highway Administration
GPS	Global Positioning System
GTFS	General transit feed specification
LEHD	Longitudinal Employer-Household Dynamics
MaaS	Mobility as a Service
MNL	Multinomial logit
MPO	Metropolitan planning organization
NHTS	National household travel survey
NL	Nested logit
NN	Neural network
NYC	New York City
NYMTC	The New York Metropolitan Transportation Council
ODRS	On-demand ride service
OLS	Ordinary least squares
RF	Random forest
RHTS	Regional household travel survey
SVM	Support vector machine
TNC	Transportation network company
VMT	Vehicle miles traveled
XGB	Extreme Gradient Boosting

SUMMARY

Recent advances in information technology have facilitated the emergence and growth of travel modes like ride-sourcing, car-sharing, and bike-sharing, providing travelers with unprecedented and broad travel options. The nature of these options will significantly affect the way people travel and engage in activities, and therefore lead to transport network impacts. Ride-sourcing, referring to app-based on-demand ride service (ODRS), exhibits some similar traits of traditional taxis but provides better real-time information, more flexible pick-up locations, normally shorter wait time, and lowered costs. Ride-sourcing can be seen as an improved service of traditional taxi, but it is also criticized by its "black boxed" pricing mechanism and its lack of drivers' strict background check or training that brings about security concerns. Ride-sourcing also share similarities with automated vehicles, as both provide request-based point-to-point mobility service. Though automated vehicles will certainly evolve into more advanced and specified mobility forms and will further transform the transportation landscape, ride-sourcing can be seen as a transitional form and litmus test for automated vehicles.

The rise of ODRS reflects the shift away from personally owned modes of transportation and towards mobility solutions that are consumed as a service, also known as Mobility as a Service (MaaS). Both ride-sourcing and taxis are publicly available, provides point-to-point service, and does not require the traveler to own vehicles, which make them quite unique compared to other public or private mobility options. Compared to traditional taxis, ride-sourcing has greater ease of use, provides better real-time information, and lowers the cost of traditional taxis, which facilitates the fast expansion of

its market shares. By providing point-to-point mobility options and easily integrating into gaps in service provided by public transit, ODRS has the potential to elevate mobility and accessibility generally and particularly for transit-dependent travelers.

In the American context, the vast difference between mode shares of driving versus alternative transportation modes has not only resulted in traffic congestion on most roads and highways, but also spawned the lack of funding for constructing infrastructure of all other modes. Rail transit, bus transit, paratransit, and bike and pedestrian infrastructure have suffered from lack of funding and investment, many of which should have become more prevalent travel options in cities, especially for those who cannot drive. Urban accessibility and mobility are impaired, on one hand, by the increasingly severe roadway traffic congestion, and on the other hand, by the lack of convenient and affordable travel options especially for physically or economically disadvantaged people. Even though ODRS are mostly available in urban and densely developed areas, the recent fast growth and improved service of ODRS reveals new possibilities to enhance transport benefits for the larger population and make transport systems more efficient and inclusive.

The broad definition of ODRS refers to any ride service that can be requested by the users, which should include but not limited to taxis, rider-sourcing, ride-sharing, jitneys, pop-up bus, paratransit, non-emergency medical transport, and even ambulances. This dissertation focuses on taxis and ride-sourcing, which are the main forms of ODRS for everyday travel. Our understanding of taxi and ride-sourcing is very limited and it is only recently that more attention has been paid to these travel modes. Ride-sourcing refers to user-initiated app-based on-demand ride service such as Uber and Lyft and these are also known as "Transportation Network Companies" (TNCs), "dynamic ride-sharing", “ride-hailing”, and other names. Many important questions about ride-sourcing from the planning perspective are waiting for answers, such as: what is the relationship between ride-sourcing and other existing travel modes; what travel demand does it serve now and in the future; how will it impact urban transport accessibility, mobility, and equity; how should cities, transportation planners, engineers, and policy-makers intervene to maximize the benefits of ODRS, while minimizing the negative impacts of ride-sourcing. The lack of understanding of these questions not only results from the lack of empirical data of ride-sourcing, but is also related to the insufficient understanding of the multimodal nature of our transportation system. Traditional taxis share many similarities with ride-sourcing and can be an important source for understanding characteristics of ODRS, but taxi trips have been understudied in previous travel and mobility research over decades at least in the U.S.

This dissertation investigates several key aspects of ride-sourcing and taxi trips, to provide original knowledge about ODRS. More specifically, the dissertation examines the relationship between ODRS and fixed-route public transportation, identifies the characteristics of ODRS riders and trips, explores methodological improvement of modeling the choice of ODRS, and forecasts the potential impact of ODRS on transport accessibility and equity. Several different data sources are used in this dissertation. These include the 2017 National Household Travel Survey (NHTS) data, the regional household travel survey data from the New York metropolitan area, the Puget Sound region, and the Delaware Valley region, and taxi and ride-sourcing trip data from New York City. Most of the publicly available data sources of ODRS trips are used in the dissertation to further the understanding of this rapidly growing mode that may shift the transportation planning paradigm in near future. The dissertation generates new knowledge not only about ODRS, but also about the multimodal characteristics of our transport systems and explores the possibility and benefits of applying machine learning to transportation planning.

Analytical results of the dissertation reveal the important role that ODRS has in serving transport-disadvantaged populations, filling gaps in the transit system, and connecting multimodal trips, particularly in urban areas. The dissertation also contributes to identifying the socio-demographic, built environment, and trip characteristics related to the choice of ODRS. There still remain an unmet need for ODRS research in small urban and rural areas to address its value to improve mobility. Results of the dissertation unveil the substantial potential accessibility and equity benefits of integrating ODRS with transit. A primary result of the dissertation is the demonstrated strong performance of machine learning-based travel mode choices and suggests further integration of machine learning with travel demand forecasting. ODRS shares key similarities with automated vehicles that will possibly become a core of future sustainable transportation systems. The findings reveal the potential of ODRS in elevating transport benefits of the existing infrastructure and point to strategies for leveraging ODRS and automated vehicles to improve transport mobility, accessibility, and equity. Starting to incorporate ODRS into normal urban and transportation planning process has become more important than ever. The results also reveal challenges of realizing the benefits of ODRS and incorporating ODRS into travel demand forecasting, which will have to rely on data collection, public-private collaboration, and research and practical exploration of synergizing ODRS with other travel modes.

The dissertation has nine chapters. The first four chapters including Introduction, Literature Review, and Research Question introduce the objectives of the dissertation and put the dissertation into a theoretical framework. Chapter 4 illustrates the methodology and data sources. Chapters 5, 6, and 7 correspond to the three primary research questions: (1) what is the role of ODRS in urban transportation; (2) how to model the choice of ODRS in a travel demand forecasting context; and (3) what is the potential impact of ODRS on transport accessibility? The analytical results and major findings are presented in each chapter. Chapter 8 integrates the results from all analytical pieces, summarizes major findings, and discusses planning and policy implications. The dissertation ends with Chapter 9 that briefly summarizes the whole dissertation and points to next steps.

CHAPTER 1. INTRODUCTION

1.1 Goals and Objectives

For years, city and transportation planning practices have been aimed on moving people out of their cars. Although there have been practical initiatives and research break-throughs about encouraging mode shift to alternative travel modes, the automobile started to take a dominant role since 1930s and the shares of driving versus transit and active modes have not changed much since the 1970s (see Figure 1.1). The vast difference between mode shares of car versus alternative transportation modes has not only resulted in traffic congestion on most American roads and highways, it has contributed to the continuing lack of funding for constructing infrastructure of other modes. Rail transit, bus transit, paratransit, bike infrastructure, and sidewalks suffered from lack of funding and investment, many of which should have become more prevalent and heavily used in cities. Urban accessibility and mobility are impaired, on one hand, by increasingly severe traffic congestion, and on the other hand, by the lack of convenient and affordable travel options especially for the physically or economically disadvantaged people.

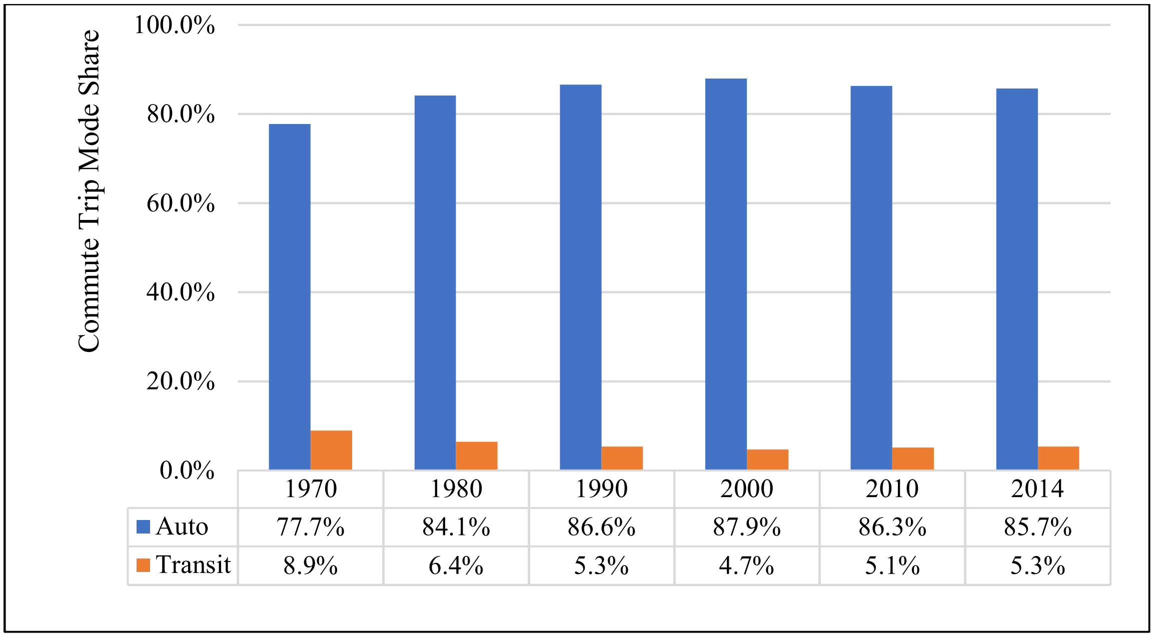


Figure 1.1. Auto vs. Transit Mode Shares in the US Since 1970
Data Source: Decennial Census Data and American Community Survey (5-Year Estimate)

Recent advances in information technologies have facilitated the emergence, availability, and growth of travel modes like ride-sourcing, car-sharing, and bike-sharing, providing travelers with unprecedented and broad travel options. The nature of these options will significantly affect the way how people travel and engage in activities, and therefore lead to transport network impacts. Ride-sourcing, referring to app-based on-demand ride service (ODRS), exhibits similar traits of traditional taxis but provides better real-time information, more flexible pick-up locations, normally shorter wait time, and lowered cost compared to traditional taxis. The fast growth of ride-sourcing can be seen as a litmus test of automated vehicles, as both of them can provide point-to-point demand based mobility service and (at least superficially) the main difference is whether there is a driver in the car. Automated vehicles will be able to further transform the transportation landscape and ride-sourcing serves as valuable evidence for researching and forecasting the impact of automated vehicles in this transitional stage. As part of the phenomenon known as Mobility as a Service (MaaS), on-demand ride service (ODRS) has been acquiring increasingly larger market shares. ODRS is publicly available, provides point-to-point service, and does not require the traveler to own any vehicle, making it unique and different from both conventional private and public mobility options. ODRS has the potential to elevate mobility and accessibility generally and for some transport-disadvantaged populations.

However, little is understood about the potential impact of ODRS. Ride-sourcing, referring to app-based on-demand ride service such as Uber and Lyft, has received great attention recently. Though ride-sourcing is often criticized for competing with public transit and luring customers with black-boxed price calculating mechanism, it also presents itself as a new form of public transportation particularly for urban areas (Rayle et al., 2016; Smart et al., 2015). Many interesting questions about ride-sourcing from the planning perspective are awaiting answers. What is the relationship between ride-sourcing and other existing travel modes? What travel demand is it serving now and in the future? How will it impact urban accessibility, mobility, and transportation equity? How should city and transportation planners intervene to maximize the benefits while minimizing the negative impacts? How can ODRS leverage existing multimodal systems to improve and expand seamless multimodal travel experience? The lack of understanding of these questions not only result from the lack of empirical data of ride-sourcing, but also echoes with the insufficiency in our understanding about the multimodal nature of our transportation system.

As a new form of ODRS, ride-sourcing is unique as it relies on real-time information provision, but it also has many similarities with traditional taxis. Traditional taxi trips are an important source for understanding the general characteristics of ODRS, including ride-sourcing. However, taxi trips have been understudied in previous travel and mobility studies. This dissertation investigates several key aspects of ride-sourcing and taxi trips, aiming to provide original knowledge about ODRS, the multimodal mechanism of our transportation system, and the potential impact of ODRS on people's travel behavior and the performance of our transportation system.

On-demand ride service may refer to any request-based mobility service. This dissertation focuses on investigating taxi service and ride-sourcing because they are the two main forms of everyday ODRS. Because of its small mode share and newness, ODRS has not received much attention in research until recently. With the rapidly growing market share of ride-sourcing, like Uber, Lyft, and Wingz in the US, and Didi-Kuaidi in China, ODRS is beginning to attract increasingly more attention in both practice and academia. As a form of point-to-point mobility service that does not require vehicle ownership, ODRS is born with similarities with both private and public transportation, but also differs from both in many ways. The uniqueness and rise of ride-sourcing reveals new potentials and challenges, destined to bring impacts on many aspects of the existing transportation system and how travelers use it.

Using different analytical approaches, including regression analysis, spatial analysis, discrete choice modeling, machine learning, scenario analysis etc., this dissertation sets out to address three interrelated research questions : (1) what is the role of ODRS in the multimodal urban transport system; (2) why do people choose ODRS and how can we model the choice of ODRS in the context of travel demand forecasting; and (3) what is the potential impact of ODRS on transport accessibility and equity.

Researching ODRS related topics has been facing the challenge of a lack of empirical data. The 2017 National Household Travel Survey (NHTS) data, containing rich information about trips made by taxi and ride-sourcing, provides an opportunity for such research. The dissertation employs most of the publicly available data sources of ODRS to further the understanding about it from different perspectives. In addition to the 2017 NHTS data, other sources include the regional household travel survey data from the New York metropolitan area, the Puget Sound region, and the Delaware Valley region, and taxi and ride-sourcing trip data from New York City. Differences and similarities between traditional taxi trips and ride-sourcing are discussed to facilitate better understanding about general ODRS trips. Investigating the research questions not only adds to the existing scarce knowledge about ODRS, but also contributes to understanding the multimodal nature of the existing transportation system that bears potential of shifting people's travel behavior in this shared mobility era.

1.2 Dissertation Outline

The dissertation has nine chapters including this Introduction chapter. In Chapter 2, several subfields of literature, including characteristics of ODRS, mode choice forecasting, and the influence of ODRS on transport performance, is reviewed and summarized to reveal existing gaps in the literature. The research question and a corresponding conceptual framework are illustrated in Chapter 3. Chapter 4 illustrates the methodology of the whole dissertation and identifies key data sources. The main research content consists of three chapters, Chapter 5, 6, and 7, corresponding to three research questions, which are presented with analytical results, findings, and conclusions. Chapter 8 summarizes the findings from Chapter 5, 6, and 7, and provides more in-depth discussions about policy and practical implications. The dissertation concludes with Chapter 9 that highlights findings, implications, and future research directions.

1.3 Research Contribution

This dissertation makes at least five principal contributions to the academic and policy literature on ride-sourcing and taxis, the multimodal characteristics of our transportation system, and the potential impacts of general on-demand ride service. First, the dissertation furthers the understanding of ODRS from different aspects by revealing the characteristics of riders and trips of ODRS and by quantifying the relationship between ODRS and public transportation. ODRS, especially ride-sourcing, is criticized for competing with public transportation, increasing congestion, misleading consumers by opaque pricing, and catering to more young, well-educated, and wealthy people. However, in addition to serving these choice users of ODRS, the dissertation revealed the fact that both taxis and ride-sourcing are disproportionately serving transport-disadvantaged population and a significant proportion of ODRS trips are serving the routes and areas with poor access to transit or are serving the first/last mile of transit. The distinct market segmentations of ODRS reveal the contrast between captive and choice users of ODRS, who may have very different travel needs in other aspects. The analytical results also show that there may be a significant mismatch between the potential demand for ODRS and its supply, which suggests that planners and policy-makers need to take a more active role in directing ODRS service to promote more equitable access. The dissertation confirms the paratransit role of ODRS, while revealing the importance of planning and policy intervention to integrate ODRS with transit to improve transport benefits generally, and for transport-disadvantaged travelers, more specifically.

Second, the dissertation is among the first attempts of modeling the choice of ODRS, which is the first step of incorporating ODRS into travel demand forecasting. Few existing travel demand forecasting models are considering ODRS, probably due to its newness and the fact that it is changing rapidly. This dissertation, using four different household travel survey datasets, explores the modeling of the choice of ODRS and identifies the socio-demographic, built environment, and trip characteristics related to the choice of ODRS. The mode choice modeling analysis further identifies what travel demand ODRS is serving and whether the choice of ODRS varies from the choice of traditional taxis. The results contribute to both our theoretical understanding and examines how we might incorporate ODRS into travel demand forecasting and other transportation planning processes.

Third, the dissertation also explores methodological innovation by applying machine learning to travel mode choice modeling. Two machine learning models, including an extreme gradient boosting model and a random forest model, are applied to modeling travel mode choices considering availability of ODRS. The models' performance is compared with a multinomial logit model. Modeling travel mode choice is a critical step in travel demand forecasting and may also face challenges and opportunities as new travel modes and data sources become available. Recent exploration of applying machine learning to transportation topics has shown the strong predictive power of machine learning, but few studies have applied machine learning to mode choice modeling. The dissertation is among the limited number of studies that use machine learning for travel mode choice modeling and has included a relatively comprehensive list of independent variables. By comparing machine learning with statistical models, the dissertation reveals the strength of machine learning and suggests the importance of integrating machine learning into travel demand forecasting to achieve more efficient and effective knowledge generation especially in such a data-abundant era.

Fourth, the last analytical piece of this dissertation focuses on estimating the impact of ODRS on transport accessibility and equity. Improving accessibility to employment and other urban amenities across all modes has been one of the primary goals of transportation planning and research. Within the US context, there is a considerable gap in the accessibility by cars and by public transportation. Although the recent literature has focused more on transit accessibility and relevant equity issues, it is a challenge to improve transit accessibility as the result of a shortage of funding and financing and mismatch between transit supply and demand. Built upon accessibility scenario development and estimation, the dissertation is the first to quantify fine-level time-sensitive accessibility benefits of ODRS considering possible variations in its level of service. The dissertation also examines how the benefits vary across population groups with different income levels. Policy and practice implications are derived about how to integrate on-demand ride service with public transit and leverage ODRS to improve accessibility and transit service equity. The dissertation discloses the substantial accessibility and equity benefits of using ODRS for short-distance trips and for fitting in gaps in transit trips. The analytical results provide insights and an analytical framework that identifies the target areas for integrating ODRS managed by private companies with the existing public transportation to improve accessibility generally. The potential elevation in accessibility is particularly beneficial for areas with poor transit coverage or where transport-disadvantaged populations reside. The dissertation will also help agencies to prepare for the future in harnessing the technological innovation in the form of evolving automated vehicles and their potential contribution to enhancing accessibility for transit riders.

Finally, though this dissertation is framed around taxis and ride-sourcing, it adds to our knowledge about shared mobility and demand-based ride service, which are different from both private and public transportation that have formed our current paradigm of travel modes in the U.S. ODRS shares many similarities with automated vehicles that will possibly become a core component of future sustainable transportation systems. Starting to incorporate ODRS and other shared mobility options into normal urban transportation planning processes has become more important than ever, as it can help leverage this mode to elevate the performance of our transportation system and can also facilitate a smooth transition to the era of automated vehicles. Based on empirical analysis and scenario forecasting, the dissertation provides important policy and practical implications about improving transport mobility, accessibility, and equity via synergizing on-demand ride service with existing mobility options like transit and active travel modes.

CHAPTER 2. LITERATURE REVIEW

Although ODRS has recently received broad interest and attention in the academia, media, and practice, there is not much published research on ODRS. For one thing, ride-sourcing is a new phenomenon and there is limited empirical data about it, so most existing studies on ride-sourcing focuses on exploring its business model and policy implications, while a few studies collected small samples of survey data. Though taxi service has existed for decades, there are only a few studies on taxis. One reason is that taxis only account for a very small modal share overall, so does not attract much research attention and has not been included in most public travel survey data. Another reason is that since taxis are run by private companies, who are in competition with other firms, there is less motivation for participating in research led by the public sector. In sum, many questions surrounding ODRS are waiting exploration, and this dissertation can by no means address all the remaining questions. The dissertation attempts to address several questions about ODRS that are most relevant to planning and can be empirically approached with currently available data sources. This literature review summarizes what we know about ODRS so far by reviewing studies mainly from three bodies of literature: (1) the characteristics of ODRS; (2) travel mode choice modeling of ODRS; and (3) the potential impact of ODRS on transport accessibility. The chapter concludes with identified literature gaps and research needs.

2.1 On-Demand Ride Service

The broad definition of ODRS refers to any ride service that can be requested by the users, which should include but not limited to taxis, rider-sourcing, ride-sharing, jitneys, pop-up bus, paratransit, non-emergency medical transport, and even ambulances. This broad classification of ODRS had not received much attention in the transportation field because of its small mode share, compared to other travel modes. However, with the prevalence of smart phones and the advancement of real-time data processing techniques, app-based on-demand ride service, known as ride-souring or Transportation Network Companies (TNCs), has become very popular in recent years. It is expected to get larger market shares in the future, which brings about questions and challenges that call for better understanding of on-demand ride service. This dissertation is concerned more about daily travel modes so it focuses mainly on taxi service and ride-sourcing.

2.1.1 Taxi as Public Transportation

The taxi is a form of public transportation and shares similarity with paratransit. Paratransit is a type of transit service that relies on small vehicles, may not work on a schedule or fixed routes, and is adept serving travelers with disabilities. In general, paratransit can encourage modal shifts away from cars, increase travel choices, enhance mobility in poor neighborhoods, and shoulder a portion of a transit systems' peak demand (Cervero, 1997). Key operational advantages of the taxicab include its low capital requirements and its ability to serve a wide range of origin-destination pairs (Austin & Zegras, 2011). However, compared to other travel modes, the taxi is frequently overlooked in the conduct of previous research (Austin & Zegras, 2011), and this is attributable to three primary reasons (King, Peters, & Daus, 2012). First, unlike conventional transit, taxicabs generally are not directly subsidized, so there is little motivation for publicly funded research to model the cost effectiveness of taxi investment (King et al., 2012). Second, taxi service is viewed too often as a luxury good for the wealthy, even though data and studies have shown that taxis actually serve more low-income groups than the rich (Altshuler, Womack, & Pucher, 1980; King et al., 2012; and Pucher & Renne, 2003). Third, taxis that serve as transit, do not have great number of riders in most American cities because people rely on driving more often (King et al., 2012). The following review summarizes the findings from existing literature on taxi trips, focusing on three specific aspects of the taxi: travel demand, trip characteristics, and taxi's relationship with fixed-route transit.

2.1.1.1 What Travel Demand Does Taxi Serve?

Evidence from existing literature suggests that taxicabs play an important role in serving the unmet demand for mobility of the low-income population. The spatial distribution of taxi trips are found to correlate significantly with the spatial distribution of transit access and low-income households (Austin & Zegras, 2011; Schaller, 2005). Kattan, de Barros, & Wirasinghe, (2010), using taxi commuter trip data in 25 Canadian cities, identified that the total number of low-income households is one of the major factors that influence work commuting trips by taxi. King & Saldarriaga, (2016) identified that taxi users who use cash have distinguishable correlation with unbanked and low-income households.

Though it is believed that taxicabs serve the low income people who often do not own vehicles, there is also concern that the relative high cost of taxicabs may preclude certain low income people. King & Saldarriaga, (2016) examined the spatial distribution of the taxi users who use cash vs. credit cards to pay in New York City. They underscored that the use of cash to pay for taxi trips is strongly associated with neighborhoods that have high shares of unbanked households, as well as with green taxicabs, which are unable to pick up trips at the airports and central business districts. With this discovery, King & Saldarriaga, (2016) noted that not having access to mainstream financial services may become a new type of discrimination in the transportation system.

The aging population reveals a need of alternative sources of transport other than personal automobile, as aging and various physical and cognitive disabilities associated with aging can seriously affect mobility that depends on driving only (Schmöcker, Quddus, Noland, & Bell, 2005). Past studies have found that the travel behavior and activity patterns of the elderly are different from those of the general population (Alsnih & Hensher, 2003; Hildebrand, 2003; Kim & Ulfarsson, 2004; Schmöcker, Quddus, Noland, & Bell, 2008; Su & Bell, 2009). Using survey data from a Taiwanese sample, Y.-C. Chang, (2013) found that elderly air passengers prefer to ask family members to drive them to the airport, while general passengers prefer to take a taxi. Their survey results also indicated that compared to “user friendly” and “convenience for storing luggage”, “safety” is the most important consideration of mode choice for the elderly in terms of trips to the airport. The elderly were found to be less likely to use public transport than private transport for trips to the airport (Y.-C. Chang, 2013).

Past studies also found that taxis serve elderly and disabled people disproportionately (S. Rosenbloom, 2003). Many areas in London provide subsidies of Taxicard that can be used to pay for taxi rides to the elderly and disabled (Schmöcker, Quddus, Noland, & Bell, 2007; Schmöcker et al., 2005). Schmöcker et al., (2005) found that elderly and disabled people generally have fewer trips and shorter distances travelled and the possession of a Taxicard is associated with more personal business trips for both the elderly and disabled groups. Modal split analysis in Nigeria showed that the elderly use taxis for 20% of their travel, a significantly larger proportion compared to other age cohorts (Ipingbemi, 2010). Taxicabs are also widely considered to be particularly adept in serving special trip purposes, such as the trip to/from a large transportation hub (railway station or airport) or a large shopping venue (Gupta, Vovsha, & Donnelly, 2008; Schaller, 2005).

2.1.1.2 What Characteristics Do Taxi Trips Have?

Taxi trips present various characteristics that might be different from both transit and automobile trips. The spatial and temporal distribution of taxi trips also vary across multiple factors, including the time of day, day of the week, season, weather, holiday, urban environment, and etc. (H. Chang, Tai, & Hsu, 2009; Kamga, Yazici, & Singhal, 2013; Phithakkitnukoon, Veloso, Bento, Biderman, & Ratti, 2010), which make it hard to characterize taxi trips or forecast taxi travel demand.

One of the empirical studies using the New York taxi trip data is by King et al., (2012a). Focusing more on policy and regulation implications, King et al., (2012a) mapped the spatial distribution patterns of the New York taxi trips and identified the asymmetrical characteristics of taxi trips, meaning that taxi trips are more often one-way than round trip so do not have symmetrical pattern between origins and destinations. Parfenov, Weeks, & Alam, (2012) explored the NYC yellow taxi GPS data and identified the seasonality of taxi trips. They found that in summer time, there are generally less taxi trips, even though the peak of tourist activity is in the summer, which is probably because the summer time allows people to choose other travel modes, such as walk and bike that can replace short-distance taxi trips (Parfenov et al., 2012). They also found that there is no remarkable difference of taxi demands between weekdays and weekends. Espín-Noboa, Lemmerich, Singer, &

Strohmaier, (2016) used the Manhattan taxi data to identify different patterns of human mobility and found that a group of taxi rides end at locations with a high density of party venues on weekend nights in Manhattan. Christoforou, Milioti, Perperidou, & Karlaftis, (2012) examined taxi travel time characteristics in the Athens metropolitan area, Greece and found that high population density in a user's residence area is related to longer travel durations of taxi trips. It was also found that elder taxi users, men, and non-regular users experience increased travel times and low-income users have shorter trips.

Large-scale events of any kind utilize taxis to accommodate increased demand (Song, Zhang, Chen, An, & Su, 2008). Taxis are highly related to specific travel purposes such as entertainment (Dongmei, Tongyan, Zhang, & Yanmei, 2009; Nutley, 2005). Trips with specific characteristics – such as the trips from or to a large transportation hub (railway station or airport) or a large shopping venue – are frequently performed by taxis (Gupta et al., 2008). Lacombe & Morency, (2016) modeled taxi trip generation using a dataset that contains 1000 GPS taxi trips in a month in Montreal, Canada. They found that variables of income, age, transit access time, number of parking spots are associated with both taxi pick-ups and drop-offs and job types affect only drop-offs. The regression model they developed has a higher R-squared (0.55) for modeling the drop-offs compared to the R-squared value of 0.33 for pick-ups, which also reveals the asymmetric nature of taxi trips.

Another reason that make taxi trips hard to predict is that the taxi trip supply fluctuates not only according to demand, because taxi drivers are under no obligation to work at any given time or about when to offer service. Kamga et al., (2013), using the GPS taxi trip data in New York City, found that the minimum taxi supply is maintained at the level at which drivers receive approximately $20 per hour (excluding tips). There is great variation of taxi demand-supply equilibrium under different weather conditions (Kamga et al., 2013). Snow conditions do not affect the hourly revenues, but when there is rainfall, drivers make more frequent and slightly shorter trips to increase their income (Kamga et al., 2013). It was also found that within existing trip-frequency and trip-distance patterns, the impact of a recently instituted taxi fare increase in NYC on hourly revenues will vary among different time of day periods, which suggested that a fare increase has the potential to alter the temporal taxi supply as well as the taxi lease rents for certain periods. A study conducted in Taipei City showed that taxi drivers were driving without passengers for about 60-73% of their operation hours because they did not know where potential customers were (H. Chang et al., 2009; Hochmair, 2015).

A majority of existing studies on taxi trips focus on modeling taxi trip generation with socio-demographic and built environment factors. Yang & Gonzales, (2014) used the NYC taxi trip data to model the relationship between taxi trip generation and a set of variables including variables of transit service, socio-demographic factors, and employment. Yang & Gonzales, (2014) found that population, median age of population residing where the taxi trips originate or end, education level, median household income level, and total jobs have significant correlation with number of taxi trip pick-ups/drop-offs at the census tract level. One thing that worth mentioning is that since the demographic and employment data are acquired from the census tracts where the taxi trip start or end, it does not reflect the demographic characteristics of the actual taxi users, but reflect the characteristics of the areas where taxi trip start or end. Qian, Zhan, & Ukkusuri, (2015) used the New York taxi trip data to explore its patterns and found that taxi trips are mostly unbalanced between origins and destinations. They also detected that land use is an important factor associating with number of taxi trips starting or ending in an area. Qian & Ukkusuri, (2015), also using the New York taxi trip data, explored the spatial pattern of taxi trip generation using the geographically weighted regression (GWR). Urban form, represented by variables of residential area and road density, and subway accessibility were found to be significantly correlated with taxi trip generation (Qian & Ukkusuri, 2015).

2.1.2 Ride-Sourcing

Ride-sourcing refers to app-based on-demand ride service, which are also known as TNCs or e-hailing companies. Ride-sourcing has become very prevalent in recent years globally, the popular ones including Uber, Lyft, Sidecar, Wingz in the US, Didi-Kuaidi in China, and Ola in India. The rapidly growing market of ride-sourcing reflects the trend known as "Mobility as a Service" (MaaS) that describes the "shift away from personally owned modes of transportation towards mobility solutions that are consumed as a service" (Zielinski, 2016). It was estimated that there were about 21.7 million American adults who used at least once ride-sourcing service in 2015 and that number was going to double in five years by 2020 ("How Much More Can Ride-Sharing Services Grow in the US?," 2016). By March 2018, the number of Uber users in the U.S. amounted to 41.8 million and Lyft has about 32 million users (Verto Analytics, 2018).

While this type of new travel mode is attracting different users from previous market segments when creating a successful business model, ride-sourcing raises a few public interest questions. Supporters view ride-sourcing as part of a suite of transport options that provides fast, flexible, and convenient mobility in urban areas. By providing an attractive alternative to driving and filling gaps in the public transit network, these services can potentially reduce auto use, auto ownership, and associated environmental impacts (Metcalfe & Warburg, 2012; Rayle et al., 2016). However, critics blame ride-sourcing for increasing congestion, competing with public transit, misleading consumers through opaque pricing practices, catering only to the young and well-to-do, and endangering public safety (Rayle et al., 2016). Moreover, ride-sourcing also brings about regulation and legal challenges that concern academia, related industries, and policy decision makers (Cetin & Deakin, 2017; Mahesh, 2015). Aarhaug, (2014) summarized the four main market segments of taxis globally and analyzed their related regulations and economics. "There seem to be agreement that some regulation is needed based on the observation that the customer is faced with a temporary monopoly supplier when hailing a taxi" (Aarhaug, 2014).

Empirical studies on ride-sourcing are very limited due to the newness of this phenomenon and the lack of publicly available data sources. The following literature review focuses on summarizing the main findings concerning trip characteristics of ride-sourcing and its comparison between conventional taxi trips, rather than regulatory and legal issues that are beyond the scope of this dissertation.

2.1.2.1 Characteristics of Ride-Sourcing Users and Trips

Dias et al., (2017) found that the users of ride-sourcing and car-sharing in the Puget Sound region tend to be young, well-educated, higher-income, employed, and residing in a higher density neighborhood. The findings are to a large extent consistent with several other studies. Rayle et al., (2016) conducted their own survey in San Francisco to explore the characteristics of ride-sourcing users and trips and compare ride-sourcing with taxis and transit from different angles. It is found that many of the ride-sourcing trips in San Francisco are for social or entertainment purposes, though their survey sample did over-represent night trips which may be the reason for augmenting the proportion of social trips. They found that ‘ease of payment’, ‘short wait time’, ‘fastest way to get there’ are the top three reasons that people use ride-sourcing service. It was also found that the ride-sourcing trips, which are over-represented by night trips, are mainly replacing the trips that would have been done by taxi (39%) and transit (33%), but there was 5% of the ride-sourcing trips from survey were taken to connect to/from transit, indicating both a competing and complementary relationship between ride-sourcing and transit and a not completely overlapped market between ride-sourcing and taxi. It was also found that the ride-sourcing users were relatively younger and better educated and the income profile of ride-sourcing users is similar to the city’s average. This study reveals important characteristics of ride-sourcing and provides useful information, but due to the limited sample size and the fact that the surveys were sample-biased towards night and social trips, some of the quantification needs more research to explore.

Z. Chen, (2015), by conducting a survey about ride-sourcing users’ attitude and travel habits in the Pittsburgh region, found that ride-sourcing users are generally younger than the typical traveler and the service is used by a higher percentage of males than females. Social recreational trips are the predominant type of trips used for ride-sourcing followed by work trips. Ride-sourcing trip lengths are shorter for all types of trips when compared to typical trip makers and vehicle occupancy rates are generally higher (Z. Chen, 2015). The increase of ride-sourcing users mostly impacted taxi and private auto usage in the Pittsburgh region (Z. Chen, 2015).

L. Chen, Mislove, & Wilson, (2015) explored the surge pricing mechanism of Uber by emulating 43 copies of the Uber smartphone app to collect pricing data in San Francisco and the Midtown area Manhattan. Their observations about Uber's surge price algorithm raised important questions about the fairness and transparency of this system. "For example, users may receive dramatically different prices due to small changes in geolocation. Furthermore, the vague, changing aspects of the algorithm impacts drivers' ability to predict fares. The black-box nature of Uber's system makes it vulnerable to exploitation by passengers or possibly by colluding groups of drivers" (L. Chen et al., 2015).

2.1.2.2 Comparison between Ride-Sourcing and Taxi

Ride-sourcing and taxis share the fundamental similarity that they provide point-to-point mobility service based on user's request. There are also some important differences between the two. Due to the lack of empirical data of both ride-sourcing users and taxi users, the socio-demographic profiles of ride-sourcing vs. taxi are still unclear, and most findings summarized above are not conclusive due to the small sample size and geographic variation. This section of the literature review focuses on comparing ride-sourcing and taxis in terms of their service characteristics rather than users' characteristics, such as travel time, convenience, cost, comfort level, etc.

Travel time can be decomposed into three types of time for ride services, including waiting time, in-vehicle travel time, and out-of-vehicle travel time. Both ride-sourcing and taxis serve point-to-point travel, there should not be significant difference in their in-vehicle travel time, but ride-sourcing is equipped with real-time GPS that directs the drivers to the shortest routing in terms of travel time based on real-time traffic conditions. Thus ride-sourcing may have shorter in-vehicle travel time compared to traditional taxis. The out-of-vehicle travel time of ride-sourcing and taxis should be similar, but ride-sourcing might have shorter out-of-vehicle travel time too. Users often have to travel to a busy street to hail a taxi, but do not need to do so for ride-sourcing.

Wait time is a main factor that differentiate ride-sourcing from traditional taxi service. Since the ride-sourcing's apps always connect the riders to their closest drivers with a centralized dispatching system, the overall waiting time of customers is, to some degree, optimized systematically depending on different fleet size. Taxi service, though equipped with GPS mostly, could be assigned to their closest customers, most of the service is hailed on road, incurring extra routing for looking customers. It is fair to assume that with the same fleet size, ride-sourcing should guarantee less waiting time compared to traditional taxis. A recent report in Australia states that the average waiting time for a taxi is about 8 minutes, while the average waiting time of UberX passengers is about 4.5 minutes (Lambert, 2016b). Although the centralized dispatching systems guarantee the optimized waiting time of ride-sourcing passengers, the actual difference in waiting time between ride-sourcing and taxi relies heavily on their fleet sizes and locations.

The second feature that makes the waiting time of ride-sourcing distinct from conventional taxi is its symmetric information provision. Ride-sourcing apps provide real-time location of both the riders and drivers, which can improve real-time information exchange. Existing literature suggests that the time travelers spend outside the transportation vehicle of choice, such as waiting time, is more onerous than the time they spend inside the vehicle in motion to their destination (Ben-Akiva & Lerman, 1985). Also real-time transit information can be useful to transit passengers for multiple reasons, as it can allow passengers to use their waiting time more productively, select which route they would want to take, or choose to select an alternative mode of transportation (Mishalani, McCord, & Wirtz, 2006). Compared to conventional taxi service, ride-sourcing has a distinct advantage in information provision by providing the real-time location to both passengers and drivers and allowing direct phone call contacts between them. Li, Xia, & Duan, (2014) examined the changes in cost of conventional taxi service before and after the application of the taxi-calling mobile apps in Shanghai. It is found that compared to traditional taxi industry, taxi-calling mobile apps have brought about several benefits, one of which is improving the symmetry of supply and demand information. In summary, ride-sourcing can be considered as providing the service with similar but often less in-vehicle, out-of-vehicle travel time and also shorter wait time compared to traditional taxis.

Even though ride-sourcing has some advantages in shortening the travel time and providing real-time information, the main reason for passengers to choose ride-sourcing over taxi is mainly related to its lowered cost (Picci, 2016). Ride-sourcing like Uber and Lyft take a similar pricing structure as conventional taxi, whose prices vary depending on what city it is, how far the travel is, and how long the trip takes ("How Much Does Uber Cost versus Taxis?," n.d.). However, the pricing mechanisms of taxis and ride-sourcing are very different. Taxis follow a fixed ordinance, often set by the municipality, that regulates how much taxi service costs per mile traveling and per minute waiting. In contrast, ride-sourcing employs a more flexible and opaque pricing mechanism that gives passengers price quotes every time the service is requested. The price may surge significantly when there is higher demand than supply, such as during peak hours. Taking Uber as an example, it was reported that on average Uber's pricing is only better than the average taxi on trips greater than $35 cost, as Uber's price surging may double the trip's cost or more during rush hours and on holidays (Meadows, 2015). Li et al., (2014) believed that compared to the traditional taxi industry, taxi-calling mobile apps in Shanghai have brought about several economic benefits such increasing the efficiency of operation, and reducing social cost. Salnikov, Lambiotte, Noulas, & Mascolo, (2015) conducted an experiment to compare the price of taking Uber and Yellow Cab in New York for the same trip. They developed a mobile app called 'OpenStreetCab' that returns either Uber or the Yellow Cab as a cheaper mode given a certain pair of origin and destination in New York. The astounding finding of this paper is that for most trips made in Manhattan New York City, Uber costs more than the Yellow Cab without considering what time of the day the trips were taken. Though this point needs to be further researched but it is true that the real cost of ride-sourcing is not transparent as it fluctuates substantially by location and by time of day.

Ride-sharing is a critical component of ride-sourcing that differentiates it from the traditional taxi. The real-time information and centralized dispatching system of ride-sourcing can facilitate more trip matching and promote more ride-sharing compared to traditional taxis. Some studies have explored the potential that ridesharing could have in affecting vehicle miles traveled (VMT) and urban land use patterns (Fagnant & Kockelman, 2014; Shaheen, Chan, & Micheaux, 2015; W. Zhang, Guhathakurta, Fang, & Zhang, 2015). Encouraging the sharing part of ride-sourcing and ODRS increases the potential to make transport more efficient and reduce the environmental impact of the transport system.

There are many other differences between ride-sourcing and taxis, though not documented in literature, but can be perceived through the experience of using both types of service. For example, ride-sourcing trips are recorded and linked to a specific driver, which provide better resources for accident management. The safety and security of travel may be perceived as improved for ride-sourcing. Nevertheless, because of the lack of strict drivers' background check and training, travelers may perceive ride-sourcing not as "official" or "safe" as traditional taxis. These nuanced differences may be changed as both ride-sourcing and taxis evolve to improve their services.

2.1.3 ODRS' Potential in Improving Multimodal Connections

2.1.3.1 Multimodal Connections

The quality of transport networks does not only depend on the quality of the network, but also on the way the nodes and links connect to the larger multimodal network (Keijer & Rietveld, 2000). Riders must get to the station or bus stop by some means of travel and then must get from the alighting stop or station to the final station and then to the final destination (Clifton & Muhs, 2012). Better understanding of the complementary modes that support transit trips is useful for the transportation modeling, infrastructure planning, urban design, and health research communities (Clifton & Muhs, 2012). However, multimodal trips, or trips that use more than one means of transportation, have historically been underrepresented in travel surveying efforts, which lead to the lack of effective empirical data to understand multimodal travel behaviors (Clifton & Muhs, 2012). It is believed that people are only willing to walk about ¼ to ½ miles to get to a transit station, so there has been increasing planning effort on promoting bike-and-ride that can broaden catchment areas of transit stations (Clifton & Muhs, 2012). Using the data of the Dutch National Travel Survey from 1994, (Keijer & Rietveld, 2000) found that the propensity to take rail by people living within 500 meters from the rail station is 20% higher than those living 500 to 1000 meters away. Non-motorized transport modes are dominant at both the home-end and the activity-end and the home-end access mode to rail presents a high share of bicycle (Keijer & Rietveld, 2000).

With the growing need to encourage people to choose alternative travel modes other than driving, there has been an increasing attention paid on the integration of bike travel and transit. Some studies suggest that bike-transit integration will increase cycling and public transit mode share, while it reduces the use of private vehicle and congestion. Only a small amount of literature has used empirical data to measure the actual needs for such integration (Bachand-Marleau, Larsen, & El-Geneidy, 2011). While demonstrating the considerable synergy between biking and transit and its merits such as the low facility cost of bike-and-ride compared to park-and-ride, Pucher & Buehler, (2009) also pointed out challenges and problems of bike-and-ride integration, such as the constraint that transit vehicles need to provide bike racks. Taylor & Mahmassani, (1996) conducted a stated-preference survey to understand people's potential mode choice of bike-and-ride. It was found that the majority of people willing to bike and ride were within 2.4 km of the transit station, whereas those between 3.2 and 4.8 km demonstrated equal preference for car and bicycle as an access mode to transit (Taylor & Mahmassani, 1996). Bachand-Marleau et al., (2011) found that trips involving access or egress by bicycle at only one end of the trip accounted for the greatest proportion of respondents who stated they would be regular bike-transit users, according to a survey conducted in Montreal. By examining the Netherland's

Central Bureau of Statistics data, Rietveld, (2000) put forward that the bicycle is a potentially attractive access mode for railways since it allows travelers to avoid waiting at bus, metro or tram stops, and it is environmental friendly, cheap to own, and requires only modest parking space near the railway station.

2.1.3.2 The Interface between On-demand Ride Service and Public Transportation

The multimodal connection between bikes and transit has received a lot of attention in literature, mainly since biking effectively enlarges the catchment area of transit compared to walking, and lowers the cost compared to (auto) park-and-ride. However, biking has its limitations as an access mode to transit. Biking is sensitive to weather, availability and condition of biking routes, and number of people traveling. It has a distance limitation and often requires the transit to have bike racks that only allow limited number of passengers to bike to transit. ODRS thus presents its potential as an effective access mode to public transportation. First, ODRS almost requires no facility renovation to existing transit infrastructure, as there is no need for parking or adding racks to transit vehicles. Second, ODRS is not sensitive to weather, routes, or number of people traveling together. Third, ODRS does not have a distance limitation and the only constraint of using ODRS may be the availability of ODRS and its cost. Though ODRS is often considered as a more expensive mode compared to biking or fixed-route transit, it has great potential in improving multimodal connection with public transportation.

Empirically, taxi trips are found to present a multi-faceted relationship with public transportation. On one hand, taxis may compete with public transportation by providing point-to-point travel service with higher comfort level and convenience. On the other hand, taxi trips also complement public transportation by serving the first/last mile of transit. Empirically, the relationship between taxi trips and public transit has been increasingly studied recently, partly due to the billions of taxi trip data that became available in recent years in New York City. Hochmair, (2015) using the NYC yellow taxi trip data, conducted spatial negative binomial regression to explore the relationship between taxi trips and other explanatory variables, including the availability of subway and bus. Hochmair, (2015) found that in some areas the number of subway/train stations is positively associated with taxi trip counts. This is consistent with previous findings by Kattan et al., (2010) that the high correlation between public transit ridership and taxi trips could be explained by the direct demand for taxi service from major transit stations. Hochmair, (2015) also found in some other area, bus availability has a negative coefficient for the taxi count model, which may indicate that bus trips in well served areas compete with taxi trips.

Kattan et al., (2010), using taxi commuter trips' data in 25 Canadian cities, identified that the total number of work trips made by public transit is one of the most important factors that influence work commuting by taxi. Using data from the Taxicab, Limousine and Paratransit Association's (TLPA) 2002 Fact Book, Schaller, (2005) developed a regression model for the number taxicabs in 118 U.S. cities and counties. Schaller, (2005) found that three factors are most strongly correlated with the number of taxi trips, including the number of workers commuting by subway, the number of no-car households, and taxi usage for airport taxi trips. Gilbert & Samuels, (1982) provided a comprehensive historical review of the taxicab industry. They see taxis as one solution to providing satisfactory mobility to certain urban areas where conventional transit is less cost-effective. They argued that public transit programs and subsidies should incorporate taxicabs to improve the multimodal connection between taxi and transit (Austin & Zegras, 2011). The complementary role of taxicabs to transit and the multimodal nature of urban travel demand behind it should garner more attention and worth further research.

2.2 Mode Choice Modeling

2.2.1 Factors Associated with Mode Choices

People's mode choices have received great attention among studies on travel variables. Extensive studies have been done to explore what factors associate with people's mode choices and the factors can be roughly categorized into three groups, including socioeconomic characteristics, built environment/land use variables, and trip characteristics. In the planning field, studies on the relationship between built environment and land use factors and people's mode choices, often studied by controlling socioeconomic variables, have been burgeoning for decades and have aroused heated discussion about whether compact development strategies can influence people's travel behaviors. This review below summarizes what factors are identified by existing literature to be significantly associated with people's travel mode choices, among which socioeconomic, built environment and land use factors, and trip characteristics are the major types.

Built environment factors have been widely identified to have an interrelationship with people's travel behaviors. Studies on the relationship between built environment and mode choices are extremely extensive and it is impossible to exhaustively summarize them all. Ewing & Cervero, (2001, 2010) developed a very comprehensive review on over 200 studies about the relationship between travel behavior and the built environment categorized the built environment factors into five categories, named as ‘the D variables, as measures of the built environment’, including density, diversity, design, destination accessibility and distance to transit. The five categories overlap in some dimensions, but intuitively and systematically capture the most important aspects of the built environment that have impact on people’s mode choices and other travel behaviors. Table 2.1 summarizes the most commonly used built environment and land use variables that are identified to be closely related to people’s travel behavior and particularly mode choice. Policy and planning have been leveraged on this positive relationship between the D variables and people’s mode choices to promote compact development and encourage the use of alternative travel modes, such as transit, walking, and biking.

Later, the widely studied correlation between the D variables and travel behaviors started to be challenged by the ‘self-selection’ argument that the observed patterns of travel behaviors of less driving, more walking/biking, and transit use might be attributed to the ‘self-selection’ effect of people who choose the residential built environment that is most consistent with their predisposed travel behaviors. Nevertheless, multiple studies that employed various research approaches have attempted to control for residential self-selection to study the impact of built environment on travel behaviors, and nearly all of studies announced statistically significant associations between the built environment and travel behaviors, including mode choices (Cao, Mokhtarian, & Handy, 2009a; Ewing & Cervero, 2010; Mokhtarian & Cao, 2008). Thus far, the positive relationship between the D variables and mode choices towards alternative travel modes is still widely accepted. However, the magnitude of the effect is important as it varies across studies and must be distinguished from the effect of ‘self-selection’.

Table 2.1. Mode Choice Related Built Environment and Socioeconomic Factors

Density and Design Variables	Diversity, Destination Accessibility, and Distance to Transit Variables	Socioeconomic Variables
Household/Housing/Parcel density	Land use mix	Employment status
Population density	Job-population balance	Age
Retail job density	Job-housing imbalance	Gender
Job density	Non-retail job-housing balance	Income levels
Residential density	Retail job-housing balance	Vehicle ownership
Business density	Job mix	Number of people owning driver's license
Retail floor area ratio	Business types in neighborhoods	Household size
Number of retail parcels	Proportion of population within 1/4 mile of store	Life stage
Intersection density	Distance to closet commercial use/center	Household typology
Pedestrian environment factor	Population centrality	Occupation
Block length/size	Distance to CBD/downtown	Housing type
Sidewalk width/length/coverage	Jobs within one mile/accessible by walking	Percent/Number of workers in the neighborhood
Path directness	Job accessibility by auto	Percent/Number of the elderly
Street connectivity	Distance to transit stop	Percent/Number of the young
Traditional neighborhood vs. New Urbanist neighborhood	Distance to light rail	
Neighborhood with retail/park	Distance to closest rail station	

Source: Revised from (Ewing & Cervero, 2010)

It is widely accepted that travelers' socioeconomic characteristics are strongly associated with their travel behaviors. The mostly used socioeconomic variables include income, vehicle ownership, household size, number of workers, age, gender, and so on (as shown in Table 2.1). Income and vehicle ownership per person are determinant on people's choice of driving and other modes. Household size and number of workers also have critical effect on the mode choices of the people from the same household and work-related trips. Not like the built environment and land use variables that are often the focus of mode choice studies in the planning academia, socioeconomic characteristics are more often taken as control variables for in those studies or the travel demand modeling process, as there is less room for policy and planning intervention. Nevertheless, some population groups do present distinct patterns of travel mode choices that attract research interests and deserve special attention of the transportation planning practice.

The elderly are often found to have a distinguishable pattern of activities and mode choices. It is found that the elderly tend to have fewer non-home activities and related travel and travel time (Golob, 2000; Kuppam & Pendyala, 2001). This might be due to the decline in mobility of the elderly, especially when the elderly who used to drive reduce or cease driving (Burkhardt, 1999; Burkhardt, Berger, Creedon, & McGavock, 1998; Carp, 1988). Although the elderly present a decline tendency of driving, it was reported by previous studies that public transit is the least preferred transportation mode among the elderly (Burkhardt et al., 1998; Sandra Rosenbloom, 2004). Driving a car or riding in a car as a passenger is the most popular mode of transportation for the elderly (Burkhardt et al., 1998; Hildebrand, 2003; Kim & Ulfarsson, 2004; Sandra Rosenbloom, 2004), which might be due to safety and security concerns.

It is also found that low-income populations have different travel behaviors than other populations. Similar to the elderly whose mobility can be impaired by ageing, the low-income population's mobility is often reduced by economic disadvantages and the limited travel options in the areas where they can afford to live. Low income households are much less likely to have a vehicle and thus have a much larger proportion of choosing alternative travel modes, especially transit and walking (Murakami & Young, 1997). It has been consistently found that public transit systems are most appealing to low-income population (Giuliano, 2005; Murakami & Young, 1997). The relationship between the low income and walk trips depends largely on local land systems and the built environment. Murakami & Young, (1997) found that low income households have a much larger proportion of people walking to work due to less vehicle ownership nationwide. Kockelman, (1997) found that low-income households were negatively associated with walk and bike trips in the Bay Area, which was possibly due to the fact that only relatively wealthy can afford to live in walkable neighborhoods in the Bay Area and exclusionary zoning in the more central area of the city might be forcing low-income households to the region's periphery. Benekohal, Michaels, Shim, & Resende, (1994) report that low-income older people have a higher propensity toward use of public transit and walking. Georggi & Pendyala, (2000) found that both the elderly and the low-income people undertake significantly fewer long-distance trips than other socioeconomic groups and those long-distance trips were more likely to be undertaken by bus and geared towards social and personal business activities.

Female travelers are also found to present distinct travel behavior compared to men, though the findings are mixed. Cervero, (2002) found that female travelers were more dependent on cars, probably because women are often undertaking both working and household activities that require trip chaining between work, shops, and child-care centers, etc. It was found that women have greater willingness to reduce car use and potential stronger preference for public transportation (Matthies, Kuhn, & Klöckner, 2002).

Female's travel mode choice might interrelate with their household typology. Schwanen & Mokhtarian, (2005) found that females in two-or-more-worker households are less likely to commute by auto, possibly due to vehicle availability across household members and females' higher acceptance of slower travel modes.

In summary, some population groups, like the elderly, low-income, women, and disabled people, often present distinct travel behaviors that are associated with their different needs and destination choices. Consequently, these populations often yield special needs in travel options. Although continuous efforts have been made to understand travel behaviors of special populations, the mobility levels of those special population groups are often found to be limited and call for further improvement strategies.

Trip characteristics also directly influence people's mode choices. Trip characteristic variables that are often included in multinomial logit (MNL) model and nested logit (NL) models for modeling travel mode choices include in-vehicle travel time, out-of-vehicle travel time/distance, and cost of the trip, which are specific to each mode included in the model, as shown in Table 2.2. In addition to trip characteristics, some key locational and socio-demographic variables may also be included in those models and Table 2.2 presents several examples of those.

Table 2.2. Common Variables Used for Travel Mode Choice Modeling

In-vehicle travel time (mode specific)
Out-of-vehicle travel time/distance (mode specific)
Cost/disposable income (mode specific)
Autos per licensed driver
Downtown workplace
Disposable income
Primary worker in the household
Government worker
Number of workers
Employment density at workplace/distance

Many later empirical studies focused on incorporating the effect of factors other than in-vehicle/out-of-vehicle travel time and cost, which have not been fully discussed or captured in original MNL or NL models for mode choice modeling. Some real-time situations are an important type of factors that influence people's mode choice spontaneously, such as the level-of-service of certain travel modes and the weather condition. Bhat (1998) formulates an MNL model structure that accommodates variations in responsiveness to level-of-service variables across individuals with different socio-economic characteristics. Empirical results indicate that the responsiveness to level-of-service variables such as congestion level and frequency of transit service varies across individuals that can be distinguished by whether the traveler is traveling alone, gender, and income variables (Bhat 1998). Weather conditions, which are mostly represented by temperature and precipitation are found to have influence on people's travel behavior and mode choices, even though there is limited number of studies that quantify the influence with empirical data, making the findings of the relationship between weather and mode choices inconclusive (Böcker, Dijst, & Prillwitz, 2013).

Trip chaining is another important travel behavior that has interrelationship with mode choices. Trip chaining represents the propensity to link a series of activities into a multi-stop tour or journey (Dissanayake & Morikawa, 2002; Shiftan, 1998). Although trip chaining can be taken as a separate travel-related choice from mode choice, it has some interrelationship with mode choice and some studies have attempted to model trip-chaining and mode choices together (Bhat, 1997; Ye, Pendyala, & Gottardi, 2007). ODRS may also have impact on trip chaining and people's mode choice for different parts of the same tour, but rarely any study has been done with that focus.

The provision of travel information has been widely identified to have effect on people's travel behavior, especially transit-related travel behavior. Chorus, Molin, & Van Wee, (2006) developed a thorough literature review on this topic and found that though the provision of travel information appears to affect transit usage, its effect on mode shift from auto to transit is mild. The effect of providing real-time information on mode shift may become stronger in longer terms, due to the learning dynamics (Chorus et al., 2006). Similarly as travel information provision that is about predictability of the service, travel time variability has also been discussed in model choice modeling empirical studies and Noland & Polak, (2002) provided a review of such literature. Travel time variability might be a result from difference in travel time from day-to-day or over the course of the day, which is independent of congestion effects that might be stable to be anticipated by travelers (Noland & Polak, 2002). For automobile travel, the effect of travel time variability on mode choices has been modeled as "extra travel time", "safety margin", or simply as disutility in the utility maximization framework of mode choice modeling (Gaver Jr, 1968; Jackson & Jucker, 1982; Knight, 1974; Noland & Polak, 2002). It has also been modeled by the scheduling choices that can be incorporated into the utility function of MNL models (Noland & Polak, 2002; Small, 1982). For public transport services with fixed time intervals, the selection of departure time takes a form of discrete intervals corresponding to the transit's time, so travel time variability can be modeled as an inherent utility of success or failure of adherence to the schedule of the fixed route service (Bates, Polak, Jones, & Cook, 2001; Noland & Small, 1995).

There are some other factors that may also influence people's travel mode choice routinely or occasionally. Attitudes and preference factors are closely related to people's travel mode choices (Cao et al., 2009a; Cao, Mokhtarian, & Handy, 2009b). Weather conditions are found to impact travel mode choices, especially active travel behaviors (Saneinejad, Roorda, & Kennedy, 2012). Trip purpose is another important factor that directly affects people's mode choice. Most of the existing mode choice studies focus on commuting trips or home-based non-work trips and some others focus on special trip purposes or traveler groups, such as students' travel (Ewing, Schroeer, & Greene, 2004; Klöckner & Friedrichsmeier, 2011; McDonald, 2008), trips to the airport (Gupta et al., 2008), shopping trips (Bhat, 1998b), long-distance trips (Limtanakool, Dijst, & Schwanen, 2006) and so on. Table 2.3 provides a summary of the built environment, social-demographic, and trip factors that have been identified by existing literature.

Table 2.3. Summary of Mode Choice Related Factors from Literature

Built environment and land use variables type	Metrics	Literature
Land use mix	Land use mix entropy index; land use diversity index; retail floor area ratio; job-housing (im)balance;	(3); (8); (9); (10); (14); (16); (17); (21);
Land use type	Dominant land use type; distance to certain land use;	(8); (17);
Road network characteristics	Intersection density; street connectivity; percent of cul-de-sac streets; path directness; % 4-way intersections;	(5); (8); (11); (16); (18); (21);
Density	Population density; population centrality; job density;	(1); (6); (10); (3); (4); (11); (14); (16); (18); (19); (21);
Transit access	Distance to nearest transit stop; % within walking distance of bus;	(6); (16)
Neighborhood type	Neighborhood with retail; pleasant for biking/walk; distance to amenities (grocery stores, park, retail);	(9); (15);
Job access	Jobs within some distance; job accessibility by a mode;	(4); (6); (10); (11);
Location in the urban area	Distance to CBD/downtown; CBD zone vs. non-CBD;	(2); (7);
Pedestrian environment	sidewalk coverage/ratio;	(3); (9); (12);
Socio-demographic and economic variable type	**Metrics**	**Literature**
Age	Age range;	(1); (6); (8); (19);
Gender	Male or female	(1); (3); (7); (8); (9);
Household typology	Household size; number/presence of elderly people; number/presence of students; number of adults	(1); (5); (7); (9); (12); (19); (20);
Vehicle ownership	Number of vehicles in the household; household vehicles per person;	(2); (3); (6); (7); (9); (12); (14); (19); (20);
Employment status	Employment status; number of workers;	(3); (9); (19);
Income	Income ranges; dummy variables indicating low-, medium-, high-income; average income;	(1); (7); (12)
Driver's license	Driver's license holding; # of driver's license in the household;	(3); (9); (12); (20);
Education	Education level;	(9);
Occupation	Occupation	(7);
Attitudes	Attitude towards active travel	(22);
Parking or fuel cost	Average fuel cost; parking cost	(8);
Trip Variables	**Metrics**	**Literature**
Level of service	in-vehicle travel time; out-of-vehicle travel time; walk/bike travel time; cost of a certain mode;	(2); (1); (7); (14);
Time of day	departure time; peak hour;	(20);
Weather	Temperature; rain; wind speed;	(19);
Activity	Activity duration; Activity start time	(13);

(1) (Bhat, 1997)
(2) (Bowman & Ben-Akiva, 2001)
(3) (Cervero, 2002)
(4) (Cervero & Duncan, 2003)
(5) (Cervero, 2007)
(6) (C. Chen, Gong, & Paaswell, 2008)
(7) (Dissanayake & Morikawa, 2010)
(8) (Frank, Bradley, Kavage, Chapman, & Lawton, 2008)
(9) (Kitamura, Mokhtarian, & Laidet, 1997)
(10) (Kockelman, 1997)
(11) (Lund, Cervero, & Wilson, 2004)
(12) (Newman & Bernardin, 2010)
(13) (Nurul Habib, 2012)
(14) (Pinjari, Pendyala, Bhat, & Waddell, 2011)
(15) (Plaut, 2005)
(16) (Rajamani, Bhat, Handy, Knaap, & Song, 2003)
(17) (Reilly & Landis, 2003)
(18) (Rodríguez & Joo, 2004)
(19) (Saneinejad et al., 2012)
(20) (Ye et al., 2007)
(21) (M. Zhang, 2004)
(22) (Cao et al., 2009a; Cao, Mokhtarian, & Handy, 2009b)

2.2.2 Travel Mode Choice Modelling Methods

Modeling travel mode choices is never an easy task as people's choice of travel modes intertwines with many different factors and most of the factors intertwine with each other. Travel mode choice has been extensively modeled with random utility maximization theory for decades. One of the most commonly applied modeling structure is the multinomial logit (MNL) model and its variations like nested logit (NL) models. Stopher, (1969) and McFadden, (1974) are the earliest contribution to applying the MNL model to travel behavior modeling. The detailed model structure and applications were thoroughly discussed in the book by Ben-Akiva & Lerman, (1985). The MNL model captures the underlying mode choice process with utility maximization assumptions that travelers are rational decision makers who are fully informed and can choose the mode that has the largest utility for them. Though there are limitations in using utility maximization to represent the mode choice process (Gärling, 1998), MNL models are effective in quantifying the effects of trip characteristics on people's mode choice. MNL model has a closed form mathematical estimation that can be easily computed. Another merit of the MNL model is that it can always replicate the shares of different classes.

Within the framework of modeling travel mode choice with random utility maximization theory, the methodology has been continuously improved through the years via improving details of the models. Early studies often model mode choices at the trip level, while recent studies start to develop tour-level travel mode choice analysis as trips of the same tour often have strong association with each other regarding mode choices (C. Chen et al., 2008; Miller, Roorda, & Carrasco, 2005). It is often found by studies that once a car is involved in a tour, all the trips of the tour will be using the car. Developing tour-based models is also closely related to the recognition of trip-chaining characteristics of some trips that often partially or directly influence mode choices (Ye et al., 2007). Bhat, (2000) introduced the random-coefficients multinomial logit model structure that assumes the coefficients of variables of a normal MNL model are random variables, to account for taste heterogeneity in mode choice modeling. Taste heterogeneity, such as taste variation in travel time saving, has also been addressed by employing mixed logit models (Hess, Bierlaire, & Polak, 2005).

Most of the literature reviewed above adopted either binary logit models, multinomial logit models, or nested logit models to estimate travel mode choices, and some other models like structural equation model and simultaneous equations model have also been applied to mode choice modeling (C. Chen et al., 2008; Scheiner & Holz-Rau, 2007). The recent advancement in computation power has brought interest in applying machine learning to transportation topics. Machine learning models are algorithms that can learn from data without relying on rules-based programming. Machine learning is rooted in computer science but its goal is similar to statistical modeling, which is to transform data into relationship and information. Karlaftis & Vlahogianni, (2011) comprehensively reviews the differences and similarities of using statistical methods versus a type of machine learning models, the neural network model, in transportation research. As they summarized, the merit of neural network model is its flexibility of dealing with complex datasets and its great predictive power, while the most significant challenge is the lack of explanation power compared to conventional statistical models. Like the merits of neural network models, most machine learning models do not have strict statistical assumptions behind the model estimation and are thus more flexible with varied data structures and variable combination. Some machine learning techniques, such as bagging and ensemble methods, are often found to perform well with unbalanced datasets when different categories are not represented equally. Bagging, also known as bootstrap aggregating, is a type of model averaging technique that can reduce the variance and help avoid overfitting issue in machine learning. Bagging is commonly applied to decision tree models. Ensemble methods refer to learning algorithms that construct a set of classifiers and then classify new data points by taking a (weighted) vote the predictions of the classifiers. Some of the widely used machine learning algorithms, such as the random forest model and the extreme gradient boosting model, are decision-tree based ensemble methods.Table 2.4 summarizes the studies that applied machine learning methods to travel mode choice modeling worldwide. As the table shows, all the studies find that machine learning models can at least achieve the similar performance of prediction compared to conventional statistical models and several studies find that the machine learning methods are significantly better. The commonly used machine learning methods in modeling travel mode choice include decision tree model (Biagioni, Szczurek, Nelson, & Mohammadian, 2008; Celikoglu, 2006; Shukla, Ma, Wickramasuriya, Huynh, & Perez, 2015; Wets, Vanhoof, Arentze, & Timmermans, 2000; Xie, Lu, & Parkany, 2003), neural network model (Hensher & Ton, 2000; Omrani, Charif, Gerber, Awasthi, & Trigano, 2013; Rao, Sikdar, Rao, & Dhingra, 1998; Shukla et al., 2015; Xie et al., 2003; Y. Zhang & Xie, 2008), support vector machine (Biagioni et al., 2008; Omrani et al., 2013; Y. Zhang & Xie, 2008), and random forest model (Sekhar, Minal, & Madhu, 2016). Most of the models' prediction power could be improved by combining other machine learning concepts such as fuzzy sets and ensemble methods (Omrani et al., 2013; Shukla et al., 2015; Vythoulkas & Koutsopoulos, 2003).

Table 2.4. Summary of Machine Learning Travel Mode Choices Studies

Reference	Application (Trip type and choice set)	Study Area	Methods	Comparison
Shukla et al. (2015)	All trips; choice among car-driver; car-passenger; transit; walk; and bicycle	Sydney	NN (fuzzy) vs. DT (fuzzy)	The use of fuzzy sets improves the performance of NN and DT
Omrani et al. (2013)	Commute trips; mode choice among car, transit, walk/bike	Luxembourg	Evidential NN (ENN) vs. NN vs. SVM	ENN is best
Zhang & Xie (2008)	Commute trips; mode choice among drive alone; shared ride with 2 people; shared ride with 3 or more; transit; bike; and walk	San Francisco	SVM vs. NN vs. MNL	SVM is best; NN and MNL are similar
Biagioni et al. (2008)	All trips; mode choice among walk; bike; auto-drive; auto-passenger; urban bus; train; suburban bus; commuter rail	Chicago	DT vs. Naïve Bayes vs. Logistic vs. SVM vs. Ensemble method	Ensemble method is best
Andrade et al. (2006)	Shopping trips; mode choice among bus, subway, and auto	Sapporo, Japan	Neuro-fuzzy MNL vs. MNL	Neuro-fuzzy MNL is better
Celikoglu	Commute trips; mode choice between car and transit	Istanbul	Radial basis function NN vs. Generalized regression NN vs. NLM	NN models are slightly better
Vythoulkas & Koutsopoulos, (2003)	Intercity travel; mode choice between rail and car	Netherlands	Neuro-fuzzy model vs. Binary logit	Fuzzy NN is slightly better
Xie et al. (2003)	Commute trips; choice among SOV; carpool; transit; bike; and walk	San Francisco	DT vs. NN vs. MNL	DT and NN are slightly better than MNL
Hensher & Ton (2000)	Commute trips; mode choice among car (no toll), car (toll), bus, and train/light rail	Six cities in Australia	NN vs. NLM	Performance of NN and NL is mixed and generally similar
Wets et al. (2000)	All trips; mode choice among car-driver; car-passenger; bike/walk; and transit	South Rotterdam region, Netherlands	DT (C4) vs. Inducing DT (CHAID) vs. MNL	Results are similar, but DT is more robust
Rao et al. (1998)	Transit-accessing trips; mode choice among walk, bus, two-wheeler, car, taxi	Mumbai, India	NN vs. MNL	NN is much better than MNL

"NN" = Neural Network; "DT" = Decision Tree; "SVM" = Support Vector Machine; "MNL" = Multinomial Logit Model; "NLM" = Nested Logit Model.

Reviewing the eleven studies shown in Table 2.4, several important problems of machine learning travel mode choices have been revealed. First, the existing studies have only included a limited number of independent variables, so limited implications are found especially for the urban transportation planning field. Most of the studies are conducted by researchers from computer or data science field, so the lack of understanding about how machine learning can be applied to transportation and planning is a common limitation. This is also related to an important shortage of machine learning, which is the difficulty of interpreting its result. Statistical models are widely used not only in forecasting but also in interpreting the relationship between independent and dependent variables which can derive policy implications. However, most machine learning models, if not all, do not allow such quantifiable interpretation of the models' result, which limits the capability of using machine learning for analyzing policy effects. Another limitation in the studies is that all the studies frame the mode choice question as a "classification question" rather than a "discrete choice question". The classification question means that the performed models only take into consideration the observed trips, so the models are actually predicting which travel mode the trip is given other variables such as travel time and distance. However, in the travel mode forecasting context, the problem needs to be framed as a "discrete choice" question, which means that the models need to predict which travel mode a traveler is likely to choose, so the models not only consider the observed trips and the chosen mode, but also need to consider the "unchosen" mode for each observed trip. Although most of the studies included in Table 2.4 shows that machine learning can at least achieve a similar level of prediction accuracy as statistical models, how well machine learning can perform is barely understood when the question is framed as a discrete choice problem. Table 2.4 covers all the machine learning studies on travel mode choice modeling, as far as the author knows, and only three of them use data from the U.S. The lack of understanding about applying machine learning to travel mode choice modeling generally and specifically in the U.S. calls for more research and exploration in this field. Moreover, as far as the author knows, no existing research examined the mode choice of ODRS, which is another major gap that this dissertation aims to fill.

2.3 ODRS and Accessibility

The emergence and expansion of ODRS is going to not only influence people's travel behaviors, but also result in network impacts to the transportation systems. Transportation systems are complex systems and the performance of which can be measured in different ways, such as traffic, mobility, and accessibility, which captures the performance of vehicle movements, passenger movements, and the ability to reach destinations, respectively (Litman, 2003). This part of the literature review focuses on discussing the potential impact of ODRS on transport accessibility and equity, which is one of the focus areas of urban and transportation planning.

Accessibility, a core concept of urban and transportation planning, has been widely studied theoretically and empirically. The definition of accessibility takes assorted forms with consideration from different perspectives. Definitions of accessibility include "the potential of opportunities for interaction" (Hansen, 1959); "the ease with which any land-use activity can be reached from a location using a particular transport system" (Dalvi & Martin, 1976); "the benefits provided by a transportation/land-use system" (Ben-Akiva & Lerman, 1979); and "the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)" (Geurs & Wee, 2004). Geurs & Wee, (2004) identified four interrelated components of accessibility: land-use, transportation, temporal, and individual. Among the four components of accessibility, ODRS will directly influence the transportation component given other factors, but in the long run, it may also influence people's travel and activity patterns that may impact the temporal and land use components of accessibility.

Transport accessibility can be measured in different ways and Table 2.5 summarizes the major categories of how accessibility can be measured. Generally, accessibility can be measured for a location, for certain infrastructure, for an individual, and for certain utility. One of commonly used accessibility measurement is location-based accessibility that reflects how many destinations can be reached from certain location such as a block group or a census tract. The dissertation focuses on examining location-based accessibility since it most directly reflects the impact of ODRS on planning-related topics.

Table 2.5. Accessibility Measures and Components

Measures	Transport component	Individual component
Infrastructure-based measures	Travel speed; vehicle-hours lost in congestion	Trip-based stratification, e.g. home-to-work, business
Location-based measures	Travel time and/or costs between locations and activities	Stratification of the population (e.g. by income, educational level)
Person-based measures	Travel time between locations of activities	Accessibility is analyzed at individual level
Utility-based measures	Travel costs between locations of activities	Utility is derived at the individual or homogeneous population group level

Accessibility is measured for a specific travel mode. For example, accessibility by car is often far better than accessibility by transit. Nassir, Hickman, Malekzadeh, & Irannezhad, (2016) develops a thorough literature review about quantifying accessibility and particularly transit accessibility. As Nassir et al., (2016) summarized, transit accessibility has been defined in a similar fashion as accessibility, with the only difference that the mode of travel is restricted to public transportation and the impedance is thus calculated based on the transit network. Estimating transit travel time is an important step in measuring transit accessibility, and most previous studies used regional travel model to estimate travel time by transit (Boarnet, Giuliano, Hou, & Shin, 2017; Welch, Gehrke, & Wang, 2016). However, these travel models' result may not be easy to obtain, which makes it a challenge for estimating transit accessibility. The recent advancement of estimating transit travel time, facilitated by the availability of General Transit Feed Specification (GTFS) data, has enabled a lot easier estimation of transit travel time (Farber, Morang, & Widener, 2014; Karner, 2018). Karner, (2018) developed time-sensitive fine-level transit accessibility and equity analysis using only publicly available data sources, including GTFS data and the US Census Bureau's Longitudinal Employer-Household Dynamics dataset.

ODRS can have substantial impact on transit accessibility. On the one hand, ODRS can serve the first/last mile of transit. Wang & Ross, (2017) found that about eight percent of the taxi trips in New York City are taken to access/egress subway stations. Though both accessibility and the first/last mile transit access have been widely studied, there is few research that integrates both. Boarnet, Giuliano, Hou, & Shin, (2017) developed a seminal study on estimating how transit station access can influence low-wage job accessibility and showed that changing the mode of access and egress to/from stations is effective at improving transit access. On the other hand, if the expense is acceptable, ODRS can be used to serve origin-to-destination trips with flexible routing. Accessibility by driving is often times better than accessibility by transit, due to the limited transit network and low level of service that make travel time by transit way longer than driving travel time in most American cities. Boarnet et al., (2017) found that low-wage job accessibility by car is almost 30 times larger than low-wage job accessibility by public transportation in the San Diego region. Since ODRS is publicly available, it provides an opportunity to integrate automotive accessibility into improving transit accessibility.

2.4 Literature Gaps and Research Needs

ODRS has received little attention until recently when ride-sourcing gains rapid growth worldwide. Traditional taxi service is a good source of understanding travel behaviors and impacts related to ODRS, but because of multiple reasons, taxi trips have also been understudied in previous research. In light of the literature review, understanding of the three bodies of literature on ODRS is limited and the findings are inconclusive: (1) some characteristics of the riders and trips of ODRS are identified, but only from small sample-sized survey and there is significant geographical variation in the data that results in mixed findings; (2) so far there has been rarely any study that incorporates ODRS into travel mode choice modeling or travel demand forecasting; and (3) the impact of ODRS on transport performance has been discussed more in recent years, but there is little empirical evidence or quantified relationships about any of the impacts.

The first research gap in the existing literature is a fundamental question concerning ODRS: what role does ODRS play in existing transport systems and what is its relationship with other existing travel modes? Existing literature has shed some light on taxi-related travel behaviors, but no systematic theoretical framework has been built concerning the relationship between taxi trips and trip by other modes. Existing literature suggests that taxi trip may serve the travel demand unmet by other modes, such as low-income, aged and disabled populations that do not own vehicles, and taxi may either support or conflict with public transportation, but the existing studies are not able to depict a comprehensive picture about taxi riders or the relationship between taxi trips and public transportation. Ride-sourcing has been found to conflict with public transportation under some conditions and replace driving under some others. There is no convincing evidence on what travel demand and population groups ride-sourcing serves, except that it serves more young and well-educated people who are more familiar with using smart phone apps. Thus, what role ODRS plays in the existing transportation system remains extremely unclear up to this point, and it calls for more research as it is anticipated to have much larger market shares and impacts in the near future.

Secondly, our understanding about travel behaviors related to ODRS is nearly barren compared to travel behaviors related to other travel modes. There are few existing studies that examines ODRS-related travel behaviors or the choice of ODRS. It is thus very challenging to integrate this travel mode into travel demand forecasting or other transportation modeling processes. ODRS will likely shift people's travel behaviors by influencing trip characteristics that may vary under different real-time conditions. ODRS may change the in-vehicle/out-of-vehicle travel time, travel cost/distance, may also influence level-of-service, people's mode choice in different weather conditions and trip chaining behaviors. It may also shift people's mode choice by providing more real-time information and predictability of the trips. Though the factors that ODRS will influence concerning people's travel behaviors could be enlisted from existing literature, little knowledge is available about to what degree ODRS is associated with those factors. Also, who are using ODRS is not clear by far and what socio-demographic characteristics are interrelated with the choice ODRS or what types of places are generating or attracting ODRS is also unclear. There have been few travel demand forecasting models that have incorporated ODRS, though it is becoming increasingly important to integrate ODRS into transportation planning processes.

Finally, the impact of ODRS on the performance of our transport systems remains unclear. The performance of transport systems can be measured by different metrics, the commonly used ones include traffic movements (VMT), mobility, accessibility, and transport equity. As an emerging travel mode and a travel mode that is built upon new technology that changes fast, ODRS is anticipated to have enlarging market shares that will inevitably have substantial network impacts on our transport systems. For example, ODRS may impact accessibility and transport equity via the three following ways:

1) ODRS can serve some travel demand that is unmet by other existing travel modes;
2) ODRS may influence the level of service on roads by influencing the congestion level and/or influencing travel time/cost between locations and activities;
3) ODRS has the potential of improving multimodal connections so it may impact accessibility and/or equity of different travel modes and reliability by facilitating seamless travel.

Although it is foreseeable that ODRS is going to impact many different aspects of our travel and the transport system, little is known about the impacts and its magnitude.

The potential influence of ODRS is important to be researched to inform future policy and regulations regarding this new travel mode.

CHAPTER 3. RESEARCH QUESTIONS AND CONCEPTUAL FRAMEWORK

3.1 Research Questions

In light of existing theories and the literature review, three important planning-related questions regarding ODRS emerge: (1) the role of ODRS in urban transportation; (2) how to incorporate ODRS into travel demand forecasting; and (3) the potential impact of ODRS on transportation systems. The three main research gaps do not constitute a complete list of the research needs regarding ODRS, but are among the most important questions that have immediate policy and practical implications and are the questions that can be approached with currently available data. This dissertation aims to provide solid evidence with available empirical data to enhance understanding and lay the foundation for future research. More specifically, this dissertation attempts to fill in the knowledge gaps of ODRS by addressing three interrelated research questions listed below.

Research Question 1: What is the role of ODRS in the multimodal transport context? ODRS is not new to our transport systems as traditional taxi service has existed for decades. However, only until recently when ride-sourcing has become more popular, ODRS has started to receive more attention in both academia and practice. It remains unclear what travel demand ODRS serves and what the relationship is between ODRS and public transportation. The dissertation attempts to address this research question by addressing the following three sub-questions:

- To what extent does ODRS compete with or complement public transportation?

- What are the characteristics of riders and trips of ODRS?
- Based on the multifaceted relationship between ODRS and public transportation, what are the policy and practical implications for improving the multimodal connection between on-demand ride service and transit?

Research Question 2: Why do people choose ODRS and how to model the choice of ODRS in the travel demand forecasting context? Understanding people's choice of ODRS is critical for understanding and forecasting ODRS-related travel behaviors and is the first step to incorporate ODRS into travel demand forecasting models. Due to the lack of empirical data that contains travelers' socio-economic profiles and ODRS travel behaviors, understanding about such topic remains barren. This dissertation uses public available travel survey data and employs both statistical and machine learning models to explore mode choice modeling of ODRS. The research question is approached by addressing the two sub-questions:

- What socio-economic, demographic, built environment, and trip characteristics are associated with people's choice of on-demand ride service?
- How to model the choice of on-demand ride service in a travel demand forecasting context?

Research Question 3: What is the potential impact of on-demand ride service on transport accessibility and equity? ODRS will inevitably impact the performance of our transport systems, but little is known about the potential direction and magnitude of the impact. This dissertation examines and forecasts the potential impact of ODRS on transport accessibility and equity. ODRS is expected to impact transport accessibility in many ways. For example, it is able to serve some travel demand that is unmet by other existing travel modes. It may influence level of service on road by influencing the congestion level and/or influence travel time/cost between locations and activities. Also, it has the potential of improving multimodal connections, so it may impact accessibility and/or equity of different travel modes. This dissertation will focus on quantifying fine-level accessibility and equity impact of ODRS, aiming to inform policy and planning practice of leveraging ODRS to improve transport benefits. Specifically, this research question can be decomposed into two sub-questions:

- What is the potential impact of ODRS on transport accessibility considering its flexibility to serve door-to-door trips and to serve the first/last mile of transit trips?
- How does the accessibility impact vary across population groups of different income levels and what are the equity implications of ODRS?

3.2 Conceptual Framework

The conceptual framework of this dissertation is derived from existing literature and the recognition of the emerging trends in transportation technology and people's changing travel behavior. With a focus on ODRS, the objective of this dissertation is to explore the three main research questions, including the "role question", the "mode choice question", and the "accessibility question". The three research questions are three different aspects of ODRS and they also form a sequential and interrelated logic flow, as shown Figure 3.1. Different number and letter combinations denote the seven important linkages that will be carefully considered and/or examined in this dissertation.

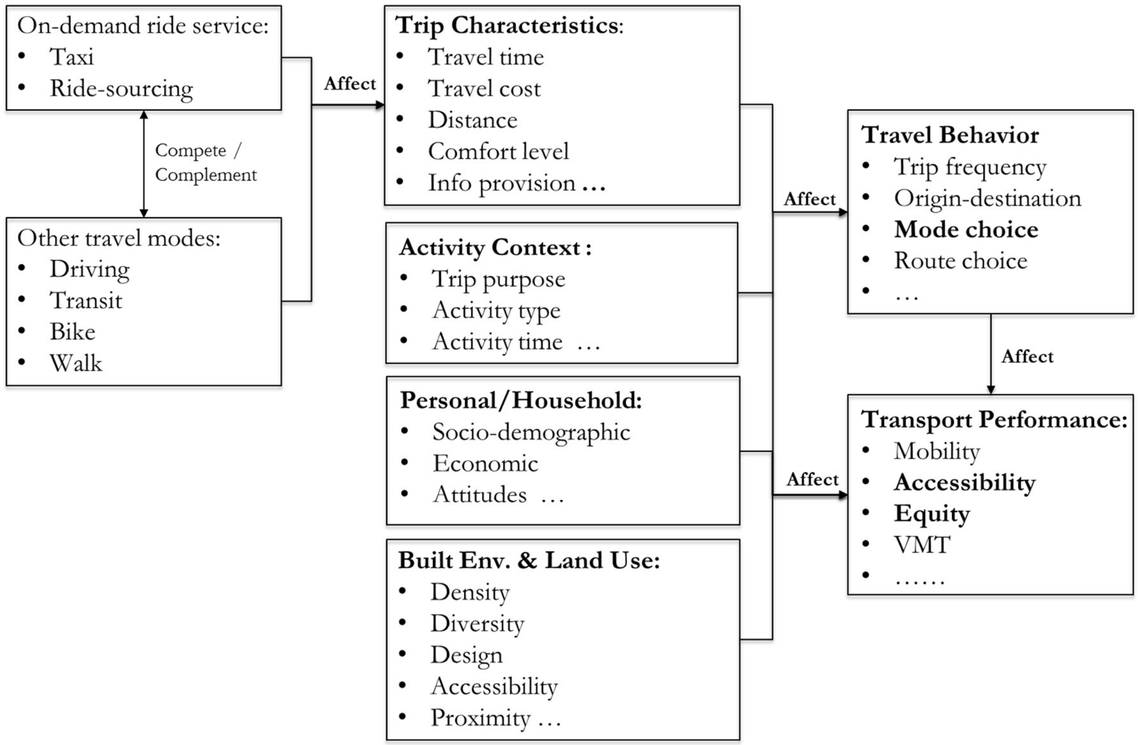


Figure 3.1. The Dissertation Conceptual Map

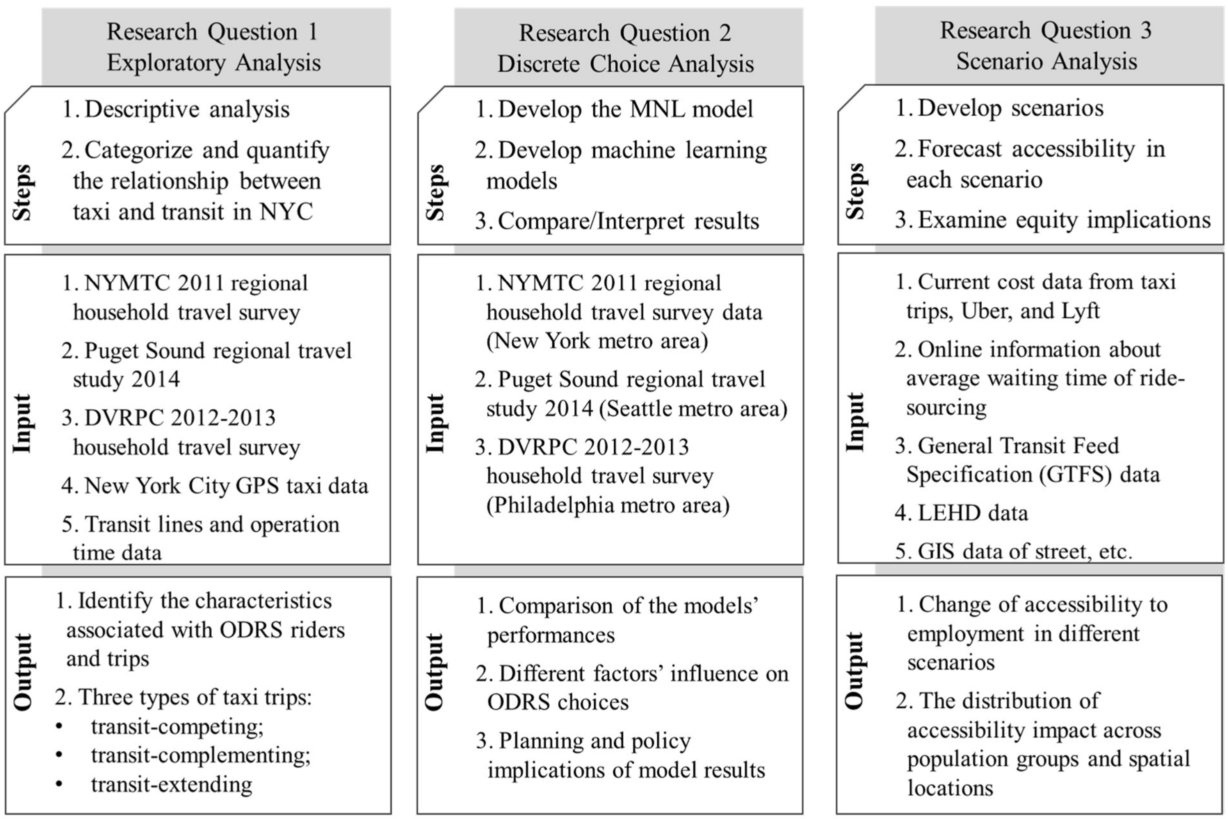
ODRS, as one of the travel modes in the transport system, may compete or complement other travel modes when serving different travel demands. The relationship between ODRS and other travel modes is the focus of Research Question 1, represented by link 1a in the conceptual map. Compared to other travel modes, a trip by ODRS may differ in waiting time, in-vehicle/out-of-vehicle travel time, travel cost/distance, comfort level, information acquisition, and other trip characteristics, which is represented by link 2a in the conceptual map. Trip characteristics, together with personal/household level factors, activity context, and built environment/land use factors, are the four main categories of factors that influence people's travel mode choices. The relationships are represented by link 2b, 2c, 2d, and 2e. Research Question 2 of this dissertation aims to explore how ODRS varies across trip characteristics compared to other travel modes, and how the choice of ODRS can be modelled given individual/household traits, trip variables, and the built environment/land use characteristics. Thus, Research Question 2 will examine links 2a to 2e. Research Question 3 examines how ODRS will impact transport performance including accessibility and equity, represented by link 3a. Accessibility is a complicated concept that interrelates with many other factors and Research Question 3 will focus on examining the direct impact of ODRS on accessibility and whether the accessibility change follows an equitable distribution across different population groups.

3.3 Methodology Overview

The dissertation methodology consists of three components corresponding to the three research questions. This section provides an overview of the methodology of the whole dissertation. Detailed methodology, implementation, and data processing procedures corresponding to each of the three research questions are illustrated in later chapters corresponding to each research question.

Data unavailability is a major challenge for conducting empirical analysis about ODRS. Currently there are few data sources that have information about ride-sourcing trips, so this dissertation relies on analyzing the limited publicly available data of ride-sourcing trips together with taxi trip data to depict a relative comprehensive picture of ODRS. Though the issue of data availability is a major limitation of the methodology of the dissertation, the dissertation aims to provide solid empirical analysis to generate new knowledge about ODRS, serve as a foundation for further research, and unveil its potential in promoting more inclusive and sustainable transportation. The data limitation and how it affects the ability to generalize the analysis are discussed in more details corresponding to each analytical step in this chapter.

The dissertation has three research questions: (1) what is the role of ODRS? (2) why do people choose ODRS and how can we model the choice of ODRS? and (3) what is the potential impact of ODRS on transport accessibility and equity? Corresponding to the three research questions, the methodology is a three-part analysis: (1) exploratory analysis of the role of ODRS; (2) discrete choice analysis of ODRS; and (3) a case study that forecasts the potential impact of ODRS on accessibility and equity (shown in Figure 3.2).



*Legend: NYC = New York City; NYMTC = New York Metropolitan Transportation Council; DVRPC = Delaware Valley Regional Planning Commission; MNL = Multinomial Logistic; LEHD = Longitudinal Employer-Household Dynamics

Figure 3.2. Dissertation Methodological Framework

The exploratory analysis of the role of ODRS integrates different data sources of taxi trips and ride-sourcing trips to extract the characteristics of ODRS riders and trips. Different data sources are used, including the GPS taxi data and ride-sourcing trip data in

New York City, GIS data of transportation infrastructure, transit operations data from General Transit Specification Feed (GTFS), Regional Household Travel Survey data from the New York metropolitan area, and the 2017 National Household Travel Survey (NHTS) data. Innovative categorization method is developed to classify taxi trips into three groups to quantify them as (1) extending, (2) competing; and (3) complementing relationships between taxi and transit. Lastly, regression analysis is developed to unravel the association of different types of ODRS trips and other variables, including socio-demographic, trip characteristics, and the built environment/land use variables. This part of the analysis aims to depict a comprehensive picture of the characteristics of ODRS trips and riders in innovative ways and reveals the role that ODRS plays in the transport system. The analytical result of the first research question lays the foundation for better understanding the second and third research questions.

The second research question is to identify what factors are associated with the choice of ODRS and how to model the choice of ODRS in the context of travel demand forecasting. Discrete choice analysis is employed for research question 2. In addition to using a multimodal logit (MNL) model, which is one of the most commonly used statistical models for travel mode choice modeling, the dissertation also employs two machine learning models, including an extreme gradient boosting (XGB) model and a random forest (RF) model. The results and performance of using the three models to predict people's travel mode choices considering ODRS are compared. Factors related to people's travel mode choices are identified. The strengths and weaknesses of using statistical models vs. machine learning are then discussed. This analysis employs the 2017 NHTS data and regional household travel survey data from three regions including the New York metropolitan area (2010/2011), the Puget Sound region (2014), and the Delaware Valley Region (2012). The new 2017 NHTS data contains rich information about trips made by taxi and ride-sourcing (Uber and Lyft) nationwide and is by far the newest and most comprehensive public travel survey data containing information about ride-sourcing trips. Regional household travel survey data from the three regions are among the limited number of regional survey datasets that have covered ODRS. Using the four different datasets to develop mode choice models of ODRS, the dissertation aims to assemble all useful information that is publicly available to explore the choice of ODRS. However, one limitation is that the current datasets do not allow an in-depth comparison between ride-sourcing and traditional taxis. The 2017 NHTS data and the regional survey data from the Puget Sound region are the only two datasets containing trips made by taxi and ride-sourcing, but ride-sourcing trips are not distinguished from trips made by traditional taxis, so the choice of ride-sourcing and taxi can only be combined in the analysis. The survey data from the New York metropolitan area and from the Delaware Valley region contain only trips made by taxi.

The third research question focuses on forecasting the impact of ODRS on transport accessibility and equity. A scenario forecasting analysis focusing on the Puget Sound region is employed. Twelve scenarios are developed for the Puget Sound region as a case study to quantify the block-group level job accessibility change due to availability of ODRS. It is assumed that ODRS can impact accessibility in two ways: 1) it can serve the first/last mile connecting to transit (multimodal travel); and 2) it can directly serve a whole trip from origin to destination (single modal trips). The twelve scenarios are developed assuming different levels of service of ODRS that are characterized by wait time of ODRS and travel distance that ODRS can be used for. Under each of the twelve scenarios, accessibility to employment is estimated and compared with the base scenario that assumes people are willing to walk up to 0.5 mile to access transit. Equity analysis is developed to compare the difference in accessibility impact across low-, medium-, and high- income workers/jobs.

CHAPTER 4. THE ROLE OF ON-DEMAND RIDE SERVICE

The role of on-demand ride service (ODRS), including taxis and ride-sourcing, is not fully understood in existing literature. ODRS has the potential to serve the travel demand unmet by other modes and may improve multimodal connectivity of the transport system. Strategies to increase transit ridership and encourage mode shift have been widely studied, and many of them are related to built environment factors and modes providing access to transit (Cervero, 1994; Cervero & Kockelman, 1997; Ewing & Cervero, 2001). Although transit ridership and people's mode choice are influenced by multiple factors, the limited coverage of transit service in many American cities discourages travelers or makes it impossible to use transit. Though new travel modes like ride-sourcing are criticized for competing with public transit, they also create potential as a new form of access to public transit (Rayle et al., 2016; Smart et al., 2015).

Little research has examined the relationship between ODRS and transit, making it difficult to understand how emerging travel modes can play a role in facilitating transit use in a multimodal context. In brief, understanding the characteristics of ODRS trips and the relationship between ODRS and transit can provide insight, not only into the multimodal mechanism of the current transportation system, but also into the potential contribution of new travel modes. Research Question 1 aims to investigate what role ODRS is playing in the transportation system by examining (1) what is the relationship between taxi and transit trips? (2) what are the characteristics of the riders and trips of ODRS? (3) what are the socio-demographic and built environment factors associated with the places that generate ODRS trips?

4.1 Methodology and Data

The methodology of Research Question 1 consists of three parts that involve using different data sources. The first part is to clarify the relationship between taxi trips and public transportation, which uses the GPS taxi trip data from New York City. By exploring the characteristics of three distinct types of taxi trips, namely the transit-competing, transit-complementing, and transit-extending trips, this part of the research reveals what travel demand taxis serve and how to integrate this mode with public transportation. The second part of is an exploratory analysis of the characteristics of riders of ODRS using the New York Metropolitan Transportation Council (NYMTC) 2010-2011 Regional Household Travel Survey (RHTS) data and the 2017 NHTS data. These two datasets contain sufficient number of trips made by taxi and/or ride-sourcing so can be used to reveal who are using ODRS. In the third analytical piece, regression models are developed to identify more clearly what the characteristics of the riders and trips of ODRS are and the places that generate different types of ODRS trips.

4.1.1 Identifying the Characteristics of On-demand Ride Service Users

Socio-demographic characteristics of riders of ODRS have not been uncovered, partly because taxi has been understudied and ride-sourcing is a new phenomenon. Also, most existing household travel survey data do not contain much data on ODRS trips, mainly because of its small shares in the market at this time. The NYMTC 2010-2011 RHTS data and the 2017 NHTS data are used to clarify the characteristics of ODRS riders and trips in the dissertation. The NYMTC RHTS data were collected during 2010 and 2011 in a 28-county area of New York, New Jersey, and Connecticut (hereafter as the "New

York region”). The dataset contains 143,925 linked trips from 18,965 households and 43,558 participants from the study area. Taxi and for-hire transportation services were included in this dataset. Conducted by the Federal Highway Administration (FHWA), the 2017 NHTS is the most recent nationwide travel survey data. It includes daily non-commercial travel by all modes, including characteristics of the people traveling, their household, and their vehicles from all members of 129,969 households nationwide, as collected from April 2016 to April 2017. The 2017 NHTS data contains more than two thousand observations of trips made by taxi and ride-sourcing, which provides the opportunity to research ODRS-related inquiries.

4.1.2 Classifying the Three Types of On-demand Ride Trips

To quantify the relationship between ODRS and fixed-route public transportation, the New York taxi trips are classified into three types: transit-extending, transit-competing, and transit-complementing trips, which are defined as follows:

1) Transit-extending trips are trips that provide connectivity to/from transit stations;
2) Transit-complementing trips serve routes and operate at times that the transit system does not serve.

Figure 4.1 is a conceptual illustration of the three types of taxi trips. Generally, the classification is based on examining whether a taxi trip could have been made by taking transit, serves the route or time that the transit system does not serve, or likely to serve as an access/egress mode to the transit. Python scripting is used to compute the relationship between the origin/destination of each taxi trip with the transit system (subway and train) and the operating time of different transit stations are also considered in the classification process. Figure 4.2 shows the stepwise logic of the classification and the detailed classification process is illustrated below.

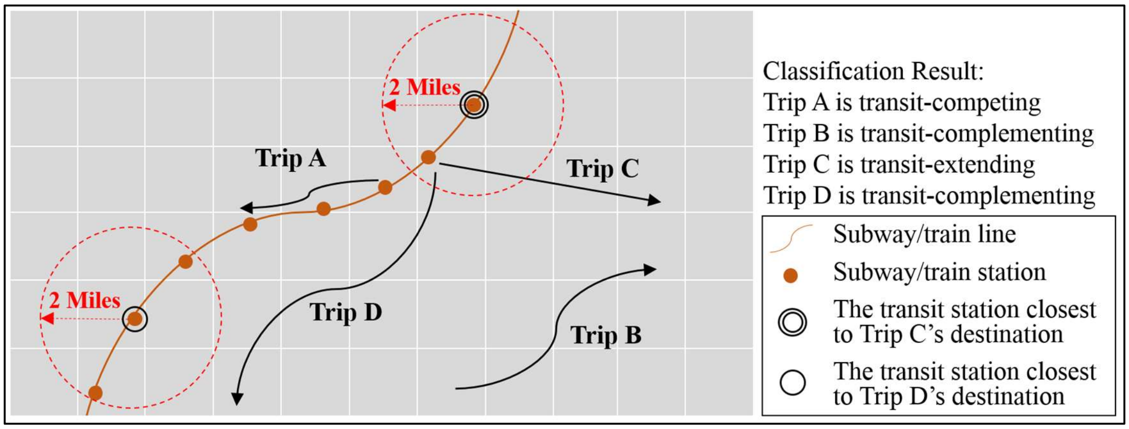


Figure 4.1. Conceptual Example of Classifying the Three Types of Taxi Trips

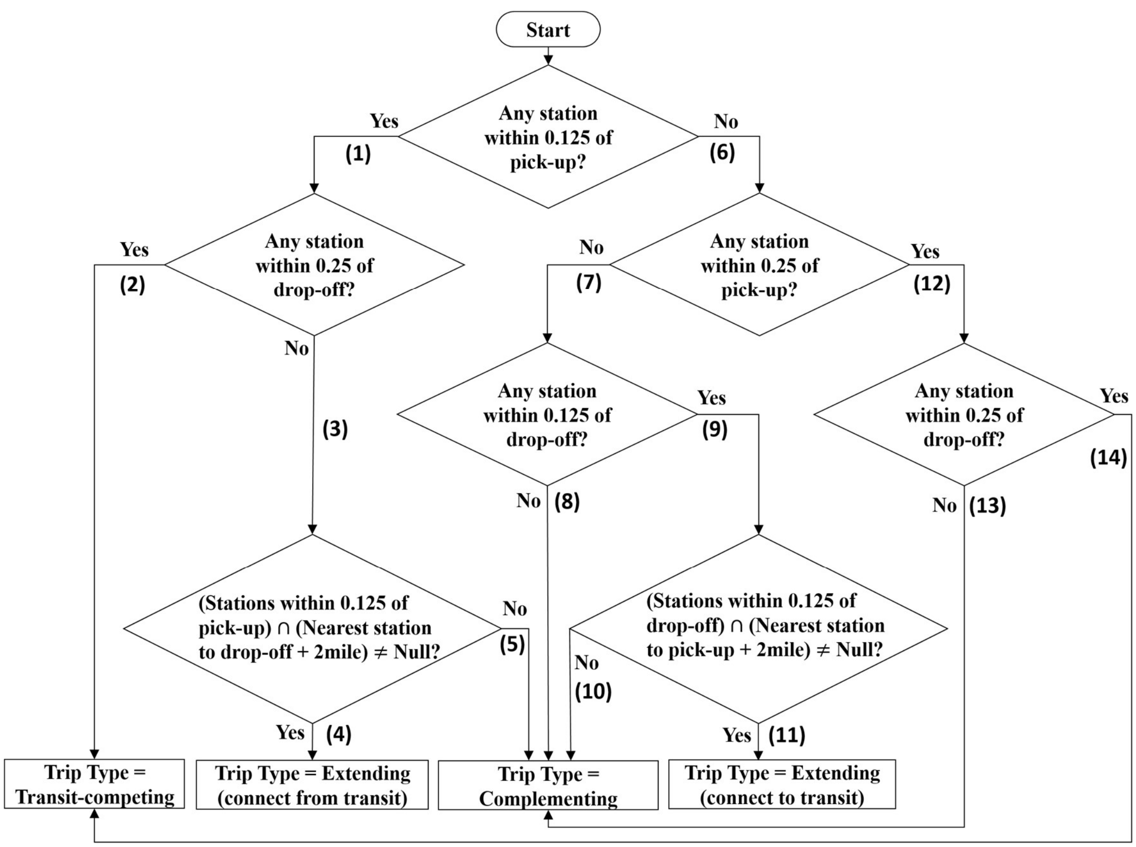


Figure 4.2. Logical Flow Chart of Classifying the Three Types of Taxi Trips

4.1.2.1 Transit-Extending Taxi Trips

Transit-extending taxi trips are defined as taxi trips that connect to/from transit. Classifying transit-extending taxi trips requires selecting trips where taxi most likely serves as the access/egress mode of transit. Transit-extending trips were selected from the taxi trip data pool according to three sequential rules: (1) only one end of the trip, either origin or destination, has at least a transit station within 0.125 mile and the trip took place when the corresponding transit service is available at that time of day; (2) the other end of the trip does not have a transit station within 0.25 mile; and (3) the transit station associated with one end of the trip is within two miles from the transit station that is nearest to the other end of the taxi trip. This third step excludes the taxi trips that have one end near a transit station, but that transit station is in the direction that conflicts the traveler's traveling direction. More specifically, if the origin of a taxi trip has a transit station within 0.125 mile and the destination of the trip does not have any transit station within 0.25 mile, the transit station located within 0.125 mile of the trip origin has to be within two miles of the transit station that is closest to the trip destination to make the trip classified as a transit-extending trip (that connects from transit). Similarly, if the destination of a taxi trip has a subway/train station within 0.125 mile and the origin of the trip does not have any transit station within 0.25 mile, the transit station located within 0.125 mile of the trip destination has to be within two miles of the transit station that is closest to the taxi trip origin to make the trip classified as a transit-extending trip (that connects to transit).

For example, as shown in Figure 4.1, Trip C starts within 0.125 mile from a transit station that is located within two miles from the station that is closest to the trip destination and the destination does not have any transit station within 0.25 miles, so Trip C is classified as transit-extending. More specifically, it is a taxi trip that connects from transit. In contrast, Trip D starts within 0.125 from a transit station, but that station does not fall into the two-mile buffer of the station that is closest to the destination of Trip D, so it is classified as transit-complementing rather than transit-extending. It is intuitive that if the rider of Trip D intends to take a taxi trip to connect from a transit station, he/she would have chosen to get off at a station that is closer to his/her destination, such as the three stations falling into the red buffer. After this final filtering step, the selected taxi trips are likely to be trips taken to connect to/from transit stations. In Figure 4.2, the logic path of (1) => (3) => (4) forms the classification process of 'transit-extending' taxi trip that connects from transit. The logic path of (6) => (7) => (9) => (11) forms the classification process of "transit-extending" taxi trip that connects to transit. The third filtering step was implemented by examining the two larger diamond boxes in Figure 4.2.

4.1.2.2 Transit-Competing Taxi Trips

Transit-competing taxi trips are defined as trips that could have been easily achieved by taking transit. The trips were selected from the taxi trip data pool by selecting the trips that have at least one subway/train station within 0.25 mile on both ends of the trip. Also, the trips had to take place when the corresponding transit service was operating. This rule of selecting transit-competing taxi trips guarantees that the same trip could have been achieved by taking transit with walking no more than 0.5 miles between the origin and destination of the trip (Trip A in Figure 4.1 is an example of such trips). Accordingly, in Figure 4.2, logic path of (1) => (2) and logic path of (6) => (12) => (14) form the classification processes of 'transit-competing' taxi trips respectively.

4.1.2.3 Transit-Complementing Taxi Trips

Transit-complementing taxi trips are defined as trips that serve the route (origin to destination) or occur at a time that the transit system does not serve. The trips were selected by excluding the previous two types of trips. Trip B in Figure 4.1 is an example of such trips. Accordingly, in Figure 4.2, four logic paths form the classification of transit-complementing taxi trips, including path that goes through links (1) => (3) => (5); links (6) => (7) => (8); links (6) => (7) => (9) => (10); and links (6) => (12) => (13).

As described, the classification of the taxi trips is based on examining the geographic location of taxi trips' origins/destinations and their relationship with transit stations' locations. The classification based on such spatial examination allows us to quantify to what degree the taxi serves as an access/egress mode of transit (transit-extending), vs. the taxi trips that are replacing transit trips (transit-competing), and the degree to which the taxi is serving trips that the transit system cannot serve efficiently. The classification method has its limitations. First, it cannot be validated with currently available data; and second, the classification cannot fully encapsulate the nature of transit-competition/complementarity. For example, if a traveler normally uses transit but takes a taxi when it is raining, such trip will be classified as transit-competing here, but it may serve as "transit-complementing" in nature since it "complements" the use of transit under certain weather conditions. However, the classification of the three types of taxi trips allows us to specify how taxis are generally utilized depending on different availability of fixed-route transit and to further understand how to maximize the synergetic relationship while minimizing the conflict between on-demand ride service and transit.

The classification is applied to the NYC GPS taxi trip data that is available on the website of NYC Taxi & Limousine Commission. The dataset contains information about every taxi trip's pick-up location, drop-off location, travel time, cost, number of passengers, trip length, etc., and includes all taxi trips from 2009 to 2016. Four-week taxi trip data in 2011 are included in our data pool, including the week of January 24th - 30th, April 4th - 10th, July 18th - 24th, and October 17th - 23rd, 2011. The four weeks were chosen because they have no major holidays and the four weeks together provide sufficient seasonal variation across the year. The 2011 data are used because ride-sourcing, like Uber, which is believed to replace taxi trips, was not available back then, so the data reflects the demand of taxis more completely. Also, the green taxi was not on the market in 2011 so there was no restriction on where a taxi could pick up passengers. Next, a random sample of size 1,000,000 trips is extracted out of this four-week data pool that contains over nine million taxi trips to allow efficient computation of different classifications and the use of regression analysis (the next step). The final classification was based on 983,053 taxi trips, after removing observations with missing or outlier values (e.g. trips longer than 100 miles were excluded). It is about ten percent of the number of trips in the original data pool and should constitute a valid sample size.

Yellow taxis were able to pick up passengers anywhere in the five boroughs of New York in 2011. Many of the taxi trips cover the eastern part of New Jersey, like Jersey City, North Bergen, and Bayonne. Therefore, the study area was set to a six-county area including New York, Bronx, Queens, Kings of New York State, and Hudson and Bergen County of New Jersey, as shown in Figure 4.3. More than 99% of the taxi trips fall into this six-county area. The subway and train lines included in the analysis are shown in

Figure 4.3, and all 1,326 subway or train stations along these lines that fall into the five-county study area were included in the classification process. Among the 1,326 stations, 899 of them are New York subway stations, 341 are along the Long Island Railroad lines, 15 are along the Metro-North Railroad lines, and 71 are along the New Jersey light rail lines. The lines and stations of the subway and trains constitute the 'transit system' in the classification of taxi trips.

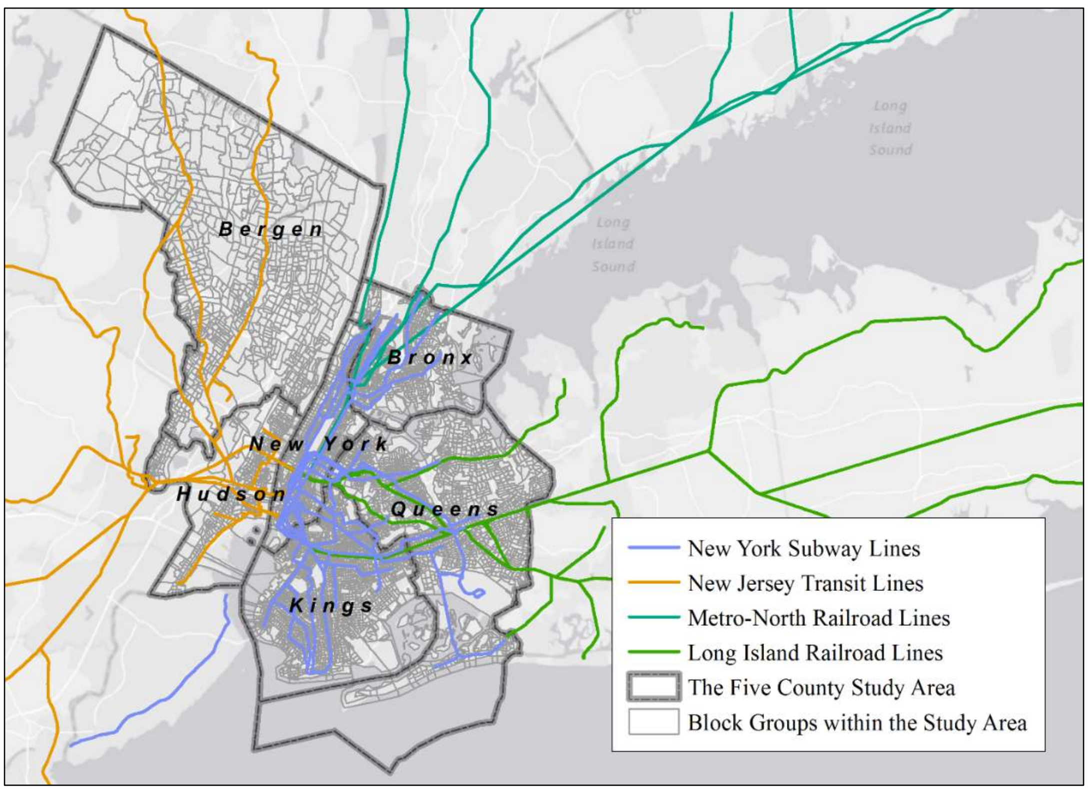


Figure 4.3. The Study Area and Transit Lines.

Bus stops were not included in the classification. There are over 22,000 bus stops in the study area. Including bus stops into the classification would result in more than 90% of the taxi trips classified as 'transit-competing' trips. However, this type of classification does not contribute much to our knowledge about the nature of the relationship between taxi and transit. Although most of taxi trips could have been made by taking the bus, given the extensive bus system in NYC, a trip that requires multiple transfers (such as transfers from subway to bus) or requires long out-of-vehicle waiting time, is not comparable to a trip made by only taking the subway or train or taxi. By examining the subway and train system in our analysis shows that all the origin-to-destination pairs of the subway system require no more than two transfers. Most of the subway stations operate 24 hours a day with short headways. Thus, the classification incorporated only subway and train lines, because whether a taxi trip is bus-competing or bus-complementing is hard to define more explicitly, given the large variation of bus headways and the extra effort of making a trip that has multiple bus transfers.

4.1.3 Regression Models

A multinomial logistic (MNL) regression model is developed to further examine the characteristics of the three types of taxi trips. The dependent variable of the model is a categorical variable indicating whether a taxi trip is transit-competing, transit-complementing, or transit-extending, given trip characteristics, socio-demographic, and built environment factors associated with the trips' origins and destinations. The variables included in the model, their data sources, and at what geographic scales they were calculated, are summarized in Table 4.1. The model is developed using the categorized taxi trip observations from the previous step. The linkage between different types of taxi trips with the economic, built environment, and trip-level factors, allows us to further understand what type of travel demand and what type of areas each of the three types of taxi trips serves.

Table 4.1. Variables Tested in the MNL Model

Variable Categories	Variable Name	Explanation	Data Source
Trip variables	Passenger	Number of passengers of the ride	NYC GPS taxi trip data
	Trip distance (log)	The log transformed trip length (in miles)	
	Payment cash	A dummy indicating whether the trip was paid by cash (vs. by card)	
	Total cost	Total cost of the trip (including fare, tip, and toll)	
	Morning peak	A dummy indicating whether the trip was between 6 am to 8 am	
	Evening peak	A dummy indicating whether the trip was between 4 pm to 6 pm	
	Late night	A dummy indicating whether the trip was between 9 pm and 4 am	
Weather variables	Precipitation	The hourly amount of precipitation when the trip took place (inch)	Weather Underground 2011 data
	Temperature	The temperature at that time (by hour) when the trip took place (°F)	
	Rain dummy	A dummy indicating whether there was rain in the hour when the trip took place	
Land use variables (at trip origin and destination)	Residential	A dummy variable of low to medium density residential land use (“one and two-family building” for New York; “Residential, single unit/low density/medium density” for New Jersey	New York and New Jersey land use and cover; Taxi trips' origin and destination are joined to the land use category that they fall into
	High residential	A dummy of high-density residential land use (“multifamily building” for New York; “Residential, high density or multiple dwelling” for New Jersey)	
	Mixed-use	A dummy of mixed use of residential and commercial	
	Commercial	A dummy of commercial land use	
	Industrial	A dummy of industrial land use	
	Other	A dummy of other land use, including vacant, open space, transportation, etc.	
Block group level variables (at trip origin and destination)	Poverty rate	Percent of families below the poverty line	ACS 5-year estimate 2009 - 2013
	Median housing value	Median housing value (in 2013 dollar)	
	Unemployment rate	Percent of unemployed population	
	Percent low-income jobs	Percent of low-income jobs (with earnings $1250/month or less)	LEHD data in 2011
	Percent high-income jobs	Percent of high-income jobs (with earnings greater than $3333/month)	
Grid level variables (at both origin and destination of the trip)	Employment density	Number of jobs per square mile	Calculated from LEHD 2010 data and New York and New Jersey roadway GIS data
	Population density	Number of persons per square mile	
	Intersection density	Number of 3-way and 4-way intersections per square mile	
	Employment-population balance	Ratio of jobs to persons	
	Bus stop density	Number of bus stops per square kilometer	New York and New Jersey transit GIS data
	Bus line density	Miles of bus lines per square kilometer	

The lack of ride-sourcing trip data has been a great challenge for conducting empirical analysis to further understand this rapidly growing travel mode. Currently, only the City of New York has publicly available ride-sourcing trip data, but the data only has the information on the pick-up location of trips made by Lyft and Uber. Uber trip pick-up location data are available for April – September 2014 in New York City and Lyft trip pick-up location data is available for 2015. To understand the difference between taxi trips and TNC trips, another two Ordinary Least Squares (OLS) regression models are developed to model the relationship between the number of taxi and TNC trip pick-ups and other socio-demographic and built environment variables of the trips' origin. In the first model, the dependent variable is the number of average daily taxi trip pick-ups aggregated at the block group level. In the second model, the dependent variable is the number of average TNC trip pick-ups aggregated at the block group level. The number of taxi trip pick-ups combines the number of yellow and green taxi trips in 2015. The number of TNC trip pick-ups covers the number of trip pick-ups made by Lyft and Uber from 2014 to 2015. The same set of independent variables in Table 4.1 are also used in this analysis to examine whether the social-demographic and built environment characteristics of places that generate taxi trips vs. ride-sourcing trips are different.

4.2 Characteristics of On-demand Ride Service Riders and Trips

The 2017 NHTS data is the most recent large-scale survey data that has information about ODRS, providing a great opportunity to understand the characteristics of the riders and trips of ODRS. There are about 2,800 ODRS trip records in the 2017 NHTS data and there is significant variation in the number of ODRS trips by states, as shown in Figure 4.4.

Most of ODRS trips concentrate in populous states. California, Georgia, North Carolina, New York, Texas, and Wisconsin have more than 100 ODRS trips in the dataset and the six states account for about 69% of all the ODRS trips in the data. About 72% of all the ODRS trips are in metropolitan areas with populations larger than one million.

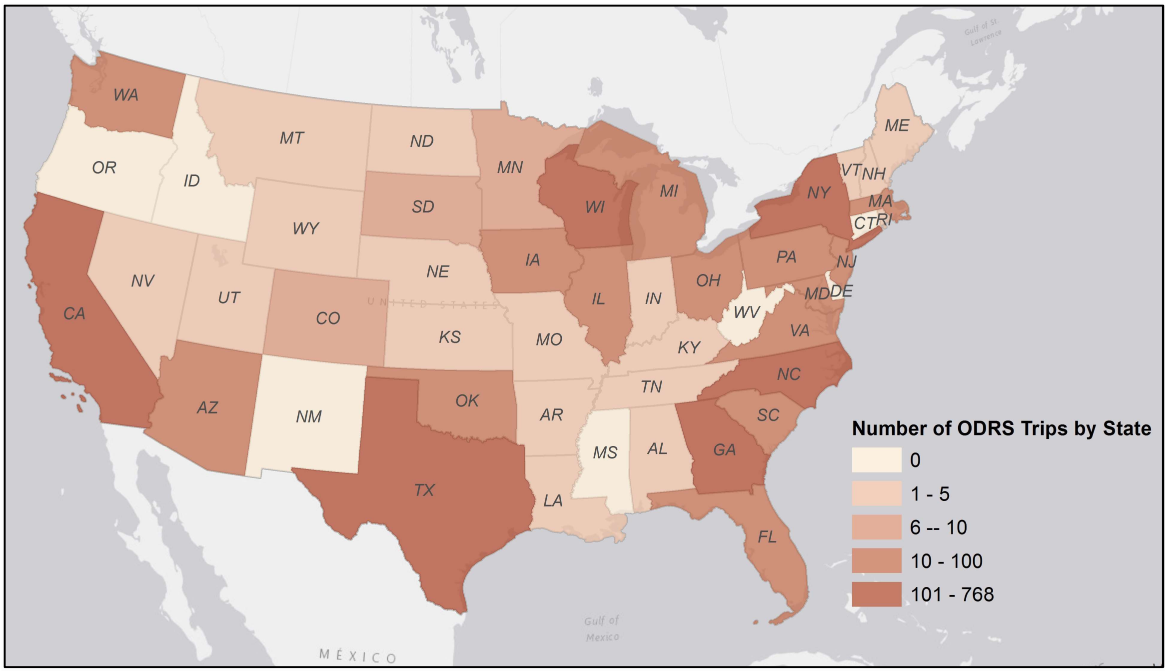


Figure 4.4. Number of ODRS Trips by State.
Data Source: 2017 NHTS Data

Some descriptive statistics of the travel time and trip distance by the five modes are tabulated in Table 4.2. The 2017 NHTS data shows that the average travel distance of ODRS (including taxi and ride-sourcing) trips is 7.9 miles which is longer than average trip length of car, biking, walking, but is shorter than that of transit. The average travel time of ODRS trips is 27 minutes including waiting time. According to the New York RHTS data that have information of trips made only by taxi, the average trip length of taxi trips is 3.8 miles, which is longer than the average trip length of biking and walking and is shorter than that of car and transit. On average, trips by taxi in the New York metropolitan area are shorter than trips made by ODRS in the national data. It may be because the higher density in NYC makes the average distance between origins and destinations shorter. It is also likely a result of the reduced cost of ride-sourcing compared to traditional taxis. Since ride-sourcing costs less per mile, people can afford and are willing to travel for longer distance by ride-sourcing compared to traditional taxis. It is also shown in the table that the travel time by ODRS is significantly shorter than the travel time by taxi (for similar trip length), which may reflect the reduced wait time of ride-sourcing compared to traditional taxis.

Table 4.2. Travel Time and Distance by Mode

	Car		Bike		Walk		Transit		ODRS*	
	Travel Time (Min)	Distance (Mile)	Travel Time (Min)	Distance (Mile)	Travel Time (Min)	Distance (Mile)	Travel Time (Min)	Distance (Mile)	Travel Time (Min)	Distance (Mile)
	The 2017 National Household Travel Survey Data									
No. of travelers	194,719		3,530		37,603		6,591		1,645	
No. of trips	773,770		7,872		79,284		13,070		2,762	
Min	1	0.0	1	0.0	1	0.0	1	0.0	1	0.1
1st Quantile	5	1.2	10	0.6	5	0.2	30	2.7	15	1.8
Median	10	2.4	15	1.2	10	0.4	45	6.1	20	3.8
3rd Quantile	15	4.6	28	2.5	20	0.8	72	15.2	30	8.8
Max	1200	801.9	504	119.6	1065	56.3	840	5315.8	505	404.0
Average	13	3.9	23	2.5	17	0.7	57	15.3	27	7.9
	The NYMTC 2011 RHTS Data (New York Region)									
No. of travelers	95,606		769		32,833		14,603		1,316	
No. of trips	97,147		783		47,106		21,399		1,334	
Min	1	0.0	1	0.0	1	0.0	1	0.0	1	0.0
1st Quantile	6	0.9	9	0.5	4	0.1	10	1.3	12	1.0
Median	13	2.4	15	0.9	5	0.2	18	2.8	19	2.0
3rd Quantile	24	6.1	25	1.7	10	0.3	31	6.8	30	4.0
Max	185	109.8	121	23.2	895	6.1	180	74.9	180	60.5
Average	18	5.3	19	1.5	9	0.2	23	6.0	25	3.8

*The 2017 NHTS data has trips made by taxi and ride-sourcing combined, and the 2011 NYMTC data only has trips made by taxi

ODRS trips are found to have a distinct pattern of departure time compared to other travel modes according to the 2017 NHTS data. As shown in Figure 4.5, there are apparently more ODRS trips in late hours and fewer ODRS trips in peak hours. People tend to use ODRS in late night hours probably because of safety considerations of other travel modes (and lack of availability of transit in many locations). The shares of ODRS in the morning and evening peak hours are much smaller compared to other travel modes, which is probably a result of the surge pricing mechanism of ride-sourcing. Figure 4.6 shows the shares of trip departure time by mode in the 2011 NYMTC RHTS data in the New York metropolitan area. The departure times of taxi trips follows a similar pattern as the ODRS trips' departure time. There are less taxi trips in peak hours and more taxi trips happen late at night.

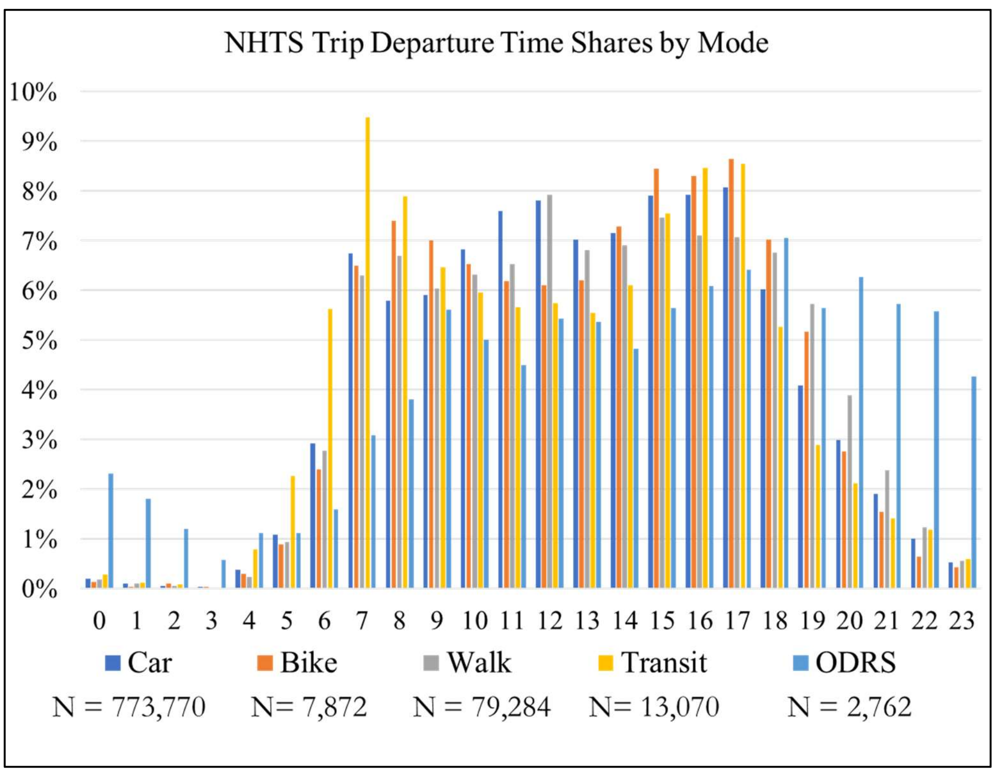


Figure 4.5. Shares of Trips by the Five Modes and by Trip Departure Time
Data Source: 2017 NHTS

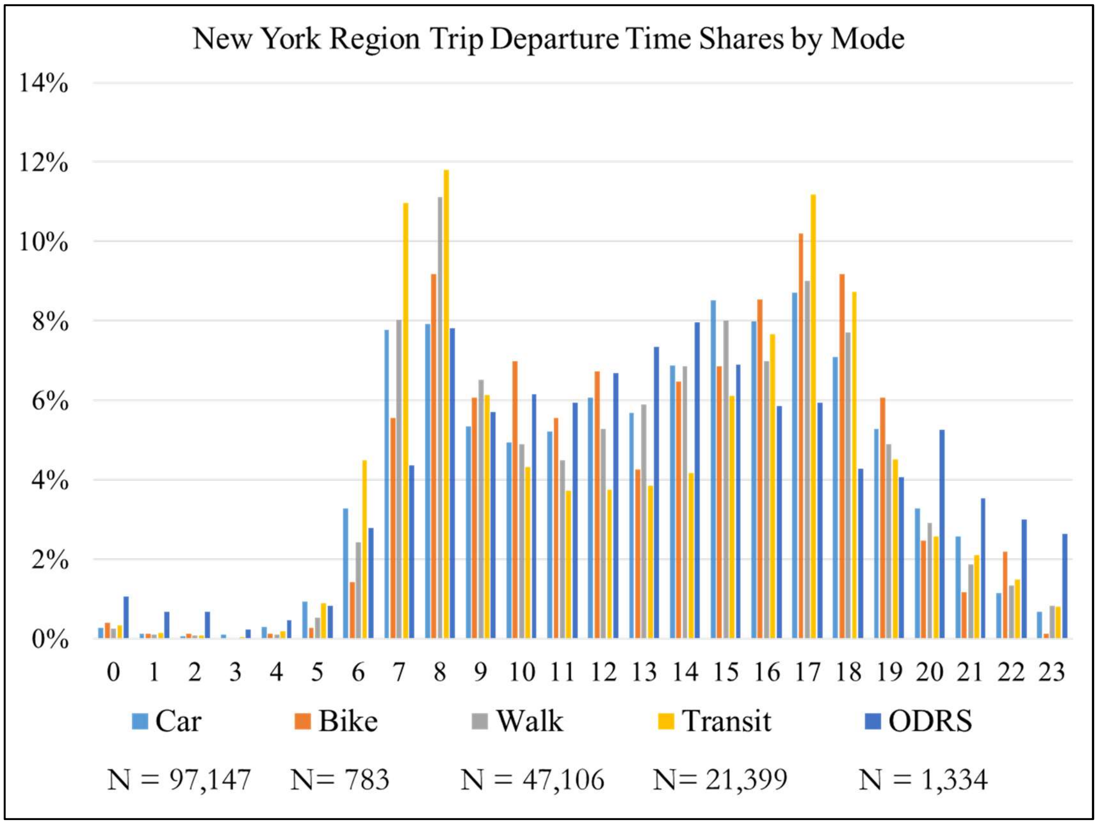


Figure 4.6. Trip Departure Time Histogram (New York Metropolitan Area)
Data Source: 2011 NYMTC RHTS Data

According to the 2010/2011 NYMTC RHTS data, the taxi has a mode share of about 1.01% in the 28 counties in the NY-NJ-CT metropolitan area. Chi-squared tests and t-tests are developed to compare the characteristics of taxi riders and travelers who choose other modes and Table 4.3 summarizes some of the socio-demographic traits of taxi riders and trip characteristics. The statistics are developed using trip weights from the original data, so it matches with the overall population distribution. Overall, taxi riders in New York consist of larger proportions of female, disabled, low-income, unemployed, and elderly/retired people, with less household vehicle ownership on average. Taxi riders' demographic characteristics are vastly different from people who drive and differ from transit riders and people who walk or bike. For example, taxis have a much larger proportion of female riders while all other travel modes have similar gender use patterns. Regarding trip purpose, taxis disproportionately serve personal/household maintenance trips that consist of trips made for personal services, appointments, and shopping needed by the individual or household. This partly explains why there is higher proportion of female riders. Previous studies have suggested the linkage between female and household-related trips and the less use of cars by females compared to males in a household (Best & Lanzendorf, 2005).

Approximately 21% of taxi riders have a disability, which is significantly larger than the disability rate of trip makers using other modes. Taxis also serve a significantly larger portion of low-income (37%) and unemployed populations (11%). About 59% of taxi riders are from households without any vehicles and the average household vehicle ownership of taxi riders is only 0.6, which is significantly lower than other travel modes. About 19.6% of taxi riders are retired, which is consistent with the previous finding that taxis disproportionately serve elderly segments of the population. In sum, 54.7% of the taxi riders are either disabled, low-income, elderly, retired or unemployed, which indicates the significant role that taxis play in providing mobility to the physically and economically disadvantaged populations.

Table 4.3. Characteristics of Taxi Riders / Trips in the New York Metropolitan Area

Demographic or Trip Factors		Taxi	Chi-Squared Test [a]	Car	Biking / Walking	Transit	Total
Number of Unique Travelers		1,316		95,606	33,602	14,603	145,127
Gender	% Female	61.3%	***	55.7%	56.2%	55.4%	55.9%
Disability	% Disabled Traveler	21.0%	***	2.7%	5.0%	4.5%	3.8%
Age	% < 16	17.1%		17.1%	17.1%	17.1%	17.1%
	% 16 - 64	70.3%		70.3%	70.3%	70.3%	70.3%
	% > 65	12.6%		12.6%	12.6%	12.6%	12.6%
Household size	Average household size	2.6	***	3.3	2.9	2.9	3.1
Travel Day	% Mon	18.8%	***	23.4%	23.4%	25.2%	23.6%
	% Tue	17.0%		21.6%	22.3%	21.2%	21.8%
	% Wed	25.3%		22.4%	22.3%	22.0%	22.3%
	% Thu	17.5%		17.6%	17.2%	17.0%	17.4%
	% Fri	21.3%		15.0%	14.9%	14.6%	14.9%
Income Levels	% Low income (< 50,000)	36.8%	***	12.5%	32.3%	30.6%	21.4%
	% Med income (50,000 - 150,000)	28.1%		44.4%	39.6%	40.3%	42.2%
	% High income (> 150,000)	35.1%		43.1%	28.0%	29.1%	36.3%
Vehicle ownership	% Zero households	59.1%	***	1.5%	48.2%	47.5%	23.0%
	Average vehicle ownership per household	0.6	***	2.1	0.8	0.8	1.5
Trip purpose	Home	38.0%	***	32.4%	24.7%	0.7%	25.9%
	Work	14.8%		12.2%	10.1%	0.8%	10.1%
	School /university	1.0%		3.3%	3.9%	0.2%	3.1%
	Escorting	4.3%		12.8%	1.8%	0.1%	7.5%
	Shopping	4.2%		11.4%	8.5%	0.3%	9.0%
	Maintenance	15.9%		10.7%	6.8%	0.2%	8.1%
	Eating out	1.7%		3.2%	3.3%	0.1%	2.8%
	Change mode / transfer	9.6%		1.6%	31.0%	97.5%	23.6%
	Airport	1.8%		0.0%	0.0%	0.0%	0.0%
	Other	8.9%		12.4%	10.0%	0.2%	10.0%
Life cycle	Student	3.6%	***	13.7%	15.7%	11.7%	14.0%
	Employed	56.0%		60.1%	58.1%	70.2%	60.7%
	Retired	15.9%		10.5%	8.2%	6.1%	9.2%
	Unemployed	11.0%		5.4%	8.4%	6.6%	6.6%
	Other	13.6%		10.3%	9.6%	5.3%	9.5%
% Transport-disadvantaged people (disability, low-income, elderly, unemployed, or retired)		54.7%	***	29.8%	42.7%	39.3%	35.4%

[a] Note: Chi-squared tests were developed by comparing taxi trips vs. non-taxi trips for all categorical variables and a t-test (assuming different variances of two samples) was developed for taxi trips vs. non-taxi trips regarding household size and vehicle ownership. '***' indicates that the null hypotheses (that taxi trips are the same as non-taxi trips regarding that variable) are rejected at the 99.9% confidence level.

The finding that approximately 54.7% of the taxi riders in the New York metropolitan area are disabled, low-income, elderly, retired or unemployed people reflects the paratransit role that taxis play. This pattern might be partially due to the 'Access-A-Ride' program that NYC launched to reduce the cost of taxis for physically challenged people, but also reflects the strong dependency of transport-disadvantaged population on taxis for other reasons, including assistance that may be provided by taxi drivers. Moreover, in the face of ride-sourcing and automated vehicles that can provide a service like taxi, but with higher level of service, it is important to consider how to improve physically or economically challenged people's access to those ODRS. NYC has announced its goal of making 50% of its taxi fleet wheelchair-accessible even though currently only about 1.8% of taxis are wheelchair-accessible (Donohue, 2013; Fitzsimmons, 2015).

Among the 923,557 unique trips that have specified travel modes in the NHTS data, 2,814 of them are made by ODRS, which is about 0.3%. A similar set of tests are developed using the 2017 NHTS data to identify the characteristics of ODRS riders and trips and the result are shown in Table 4.4. The characteristics of ODRS riders and trips revealed by the national data are different from the taxi riders' and trips' characteristics to some degree. Regarding trip characteristics, ODRS has the highest proportion of trips made for medical and dental services across the four main travel modes. ODRS also serve relatively more trips to home, work trips, and social and recreational trips. ODRS is used least for trips for school/daycare/religious activities and trips for transporting someone. Almost no ODRS trips are made for loop trips. In terms of travel day, ODRS trips have the lowest proportion of traveling on Monday and relatively lower proportion of traveling on Sunday. Traveling by car and bike or walk are spread out very evenly on the seven days of a week, while traveling by transit has much lower trips happening during the weekend. The reason why ODRS has less trips on Monday is unclear, but could be associated with flexible work schedules or long weekends of certain ODRS users.

Table 4.4. Characteristics of ODRS Riders and Trips Nationwide.

Demographic or Trip Factors		ODRS	Chi-square Tests [a]	Car	Bike/Walk	Transit	Total
Number of Unique Travelers		1,645		194,719	41,133	6,591	244,088
Trip Purpose	Home	38%	***	34%	41%	37%	35%
	Work	19%		13%	9%	22%	13%
	School/ Daycare/ Religious activity	2%		5%	6%	7%	5%
	Medical/Dental services	4%		2%	1%	4%	2%
	Shopping/Errands	8%		19%	13%	12%	18%
	Social/Recreational	15%		10%	19%	10%	11%
	Transport someone	3%		8%	3%	2%	7%
	Meals	6%		8%	8%	4%	8%
	Something else	6%		2%	3%	4%	2%
Loop Trip		0.0%	***	0.2%	14.4%	0.2%	1.9%
Travel Day	Sunday	11%	***	13%	12%	8%	13%
	Monday	7%		14%	15%	13%	14%
	Tuesday	12%		14%	15%	19%	15%
	Wednesday	15%		15%	14%	17%	15%
	Thursday	22%		15%	15%	17%	15%
	Friday	18%		15%	15%	17%	15%
	Saturday	15%		15%	13%	9%	14%
Average Household Size		2.5	***	3.1	2.9	2.5	3.1
Vehicle Ownership	% Zero Vehicle Households	32%	***	1%	20%	43%	5%
	Average Household Vehicle Count	1.3	***	2.3	1.6	1.0	2.2
Household Income	% Low-income	20%	***	15%	25%	36%	17%
	% Mid-income	42%		52%	43%	37%	51%
	% High-income	38%		33%	32%	27%	32%
Life Cycle	One adult, no children	26%	***	8%	14%	21%	9%
	2+ adults, no children	35%		22%	25%	29%	22%
	one adult with children	4%		6%	7%	9%	6%
	2+ adults with children	26%		46%	38%	27%	44%
	Retired, no children	9%		19%	16%	15%	18%

(Table 4.4 Continued)

Population Density in home location	% Low-density area (< 500 people per sqml)	8%	***	25%	12%	3%	23%
	% Mid-density area (500 - 10000 people per sqml)	43%		66%	58%	44%	65%
	% High-density area (> 10000 people per sqml)	49%		9%	30%	53%	13%
Age	% Under 18 years old	5%	***	13%	16%	8%	13%
	% 19 - 64 years old	86%		70%	71%	80%	71%
	% 65 years old or above	9%		15%	13%	12%	15%
Education	% Equal or less than high school graduate	28%	***	27%	26%	32%	27%
	% Some college or bachelor's degree	48%		54%	49%	45%	53%
	% Graduate degree	23%		19%	25%	23%	20%
% Female		48%	***	52%	50%	51%	52%
% Is working		78%	***	68%	60%	63%	67%
Use Medical Device (Disability)		11%	***	6%	6%	14%	6%
Use smart phone everyday		84%	***	77%	76%	75%	77%

[a] Note: Chi-squared tests were developed by comparing ODRS trips vs. non-ODRS trips for all categorical variables and a t-test (assuming different variances of two samples) was developed for ODRS trips vs. non-ODRS trips regarding household vehicle ownership, household size, and age. '***' indicates that the corresponding null hypotheses (that ODRS trips are the same as non-ODRS trips regarding that variable) are rejected at the 99.9% confidence level.

ODRS riders have a distinct socio-demographic profile. ODRS riders have the lowest average household size which is the same as transit riders. ODRS riders also have lower vehicle ownership on average, but it is a little higher than that of transit riders. About 32.5% of ODRS riders and about 43.4% transit riders are from zero vehicle households. ODRS riders have a slightly higher proportion of low-income families which is about 20%, but the high-income families of ODRS riders are also significantly higher than the average. This finding is like the income composition of taxi riders in New York, as it is found that taxis serve both low-income and high-income travelers disproportionately. The higher proportions of both low-income and high-income travelers in ODRS riders implies that it serves both captive and choice users, which may have very different travel needs. It is also found that over 60% of ODRS riders are from households with no children, indicating that ODRS tends to serve more people in relatively early or relatively late life stages. Significantly more ODRS trips are made in mid-density to high-density regions, which might be associated with the fact that ODRS, especially ride-sourcing, is only available in those areas. About 11% of ODRS riders use a medical device (indicating disability), which is slightly lower than the proportion of transit riders with a disability, but is significantly higher than the average. About 84% ODRS riders use smartphones every day, which is significantly higher than the proportions of smartphone users choosing other travel modes.

In the New York region, it is found that taxi serves significantly more elderly people, but this pattern is not found using the national survey data, as ODRS is found to serve less elderly people in the national data. This might indicate the difference of people using taxi vs. people using ride-sourcing. The New York data only has taxi trips in it, while the 2017 NHTS data covers both taxi trips and ride-sourcing trips. Currently, ride-sourcing trips are only available on smartphones and there might be some social and perceptual barriers for elderly people to use ride-sourcing service. Also, the New York City provides financial incentives for physically challenged people to use taxi as paratransit, which could be another reason for this difference in age cohorts of ODRS riders. The NHTS data also reveals that there are significantly more well-educated people and workers using ODRS compared to using other travel modes. Significantly less female travelers are using ODRS, which is also contradicting the finding from the New York taxi data. It is unclear why on average, less female travelers are using ODRS nationally compared to more female taxi riders in the New York region, but it could also be related to the difference between taxi and ride-sourcing and the social and perceptual barriers of using ride-sourcing, especially for certain population groups.

4.3 The Relationship between On-demand Ride Service and Transit

4.3.1 The Relationship between Taxis and Transit

Out of the 983,053 taxi trips for a month in 2011 in New York City (NYC), it is found that about 58.54% are transit-competing trips, 33.82% are transit-complementing trips, and about 7.64% are transit-extending. It is fair to say that taxis in NYC have a multi-faceted relationship with fixed-route public transit. There is great variation regarding trip characteristics across the three types, as shown in Table 4.5. Trip distance is an important characteristic that distinguishes the three types of taxi trips. The average trip length of transit-extending taxi trips is only about 1.2 miles, compared to 2.1 miles of transit-competing trips and 4.1 miles of transit-complementing trips. Regarding payment method, transit-extending trips have a much lower rate (39.6%) of paying with cards, compared to 43.6% of transit-competing and 46.4% of transit-complementing trips respectively. There is not a great variation in the number of passengers.

Table 4.5. Trip Characteristics of the Three Types of Taxi Trips

		Transit-Competing Trips	Transit-Complementing Trips	Transit-Extending Trips
No. of Observations		575,486 (58.54%)	332,531 (33.82%)	75,065 (7.64%)
No. of Passengers	mean	1.66	1.66	1.64
	sd	1.25	1.25	1.24
	min	1.00	1.00	1.00
	max	6.00	6.00	6.00
Trip distance	mean	2.09	4.14	1.19
	sd	1.81	4.36	0.63
	min	0.10	0.10	0.10
	max	81.00	82.70	41.00
Total cost	mean	10.32	15.85	7.82
	sd	5.29	12.16	2.63
	min	2.50	2.50	2.50
	max	192.90	230.00	106.20
Payment method*	Card	250801 (43.6%)	154241 (46.4%)	29724 (39.6%)
	Cash	324685 (56.4%)	178290 (53.6%)	45341 (60.4%)

*Categorical variable

4.3.2 Trip Length

The average trip lengths of the three types of taxi trips vary significantly and the distributions are shown in Figure 4.7. Overall, the clear majority (about 95.6%) of all taxi trips are shorter than 10 miles. For the transit-competing taxi trips, about 75% are shorter than 2.6 miles, and the average is about 2.1 miles, indicating that taxi service might be cheaper and fast enough over short travel distances and capable of attracting people to replace transit. Transit-complementing trips present a longer tail towards 20 miles of trip distance, reflecting the larger variation in trip length. The average length of transit-complementing trips is about 4.1 miles, which is statistically longer than the other two types of taxi trips, based on the t-test. Transit-extending trips have the shortest average trip length of about 1.2 miles, and about 90.6% of all the transit-extending taxi trips are between

0.5 miles to 2 miles long. This implies that most of the trips taken to access transit stations in NYC are beyond 0.5 mile, which is the commonly accepted walking distance for accessing transit, and shorter than 2 miles. The substantial difference in average trip lengths across the three types of taxi trips suggests that taxis are serving very different travel demands that may either extend, compete with, or complement transit services.

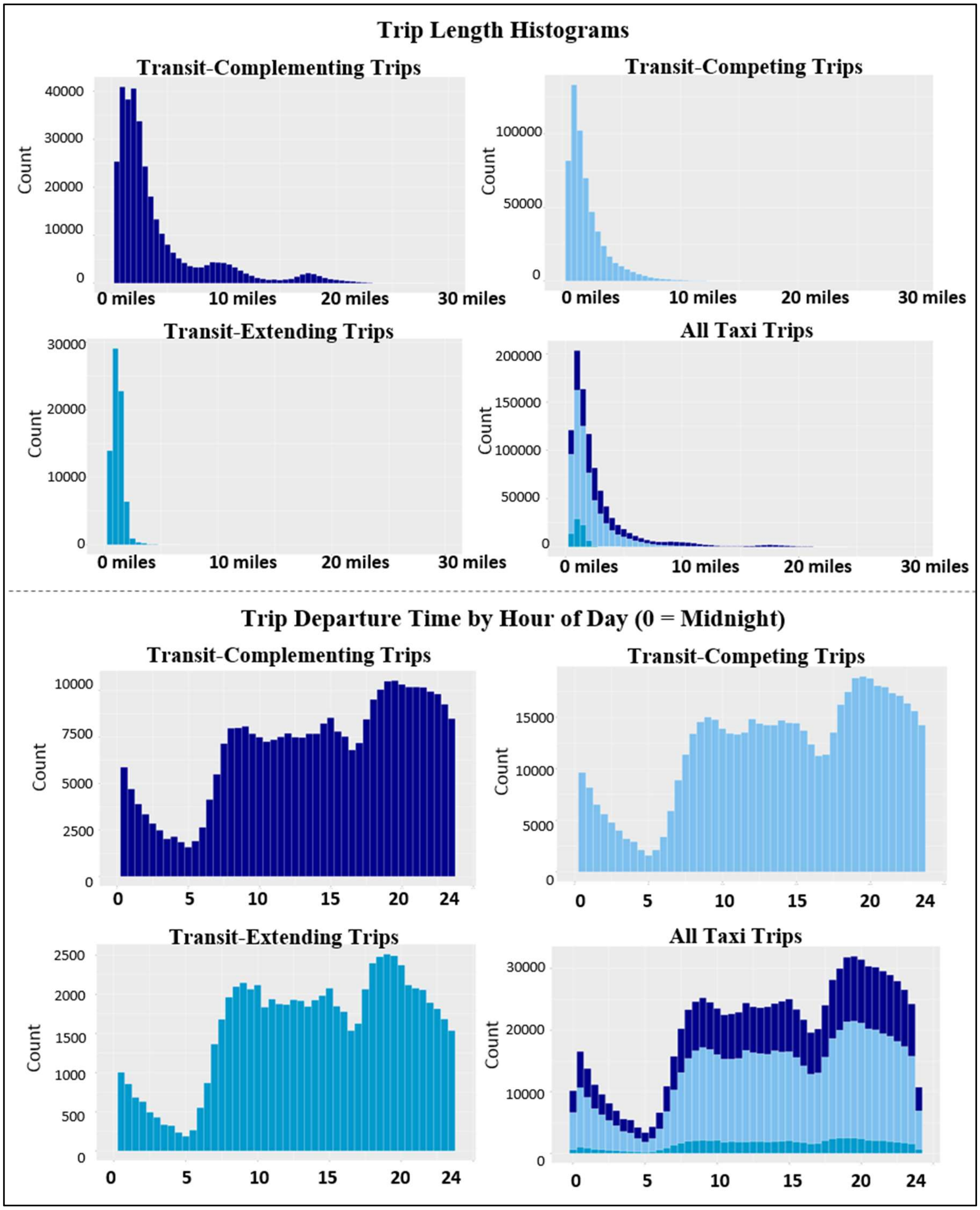


Figure 4.7. Trip Length and Time of Day Histograms of Taxi Trips

4.3.3 *Time of Day*

Figure 4.7 also presents the distributions of the taxi trips' departure time. Overall, the taxi trips in the study area generally occur during the daytime between 7 a.m. to midnight, peaking around 5 p.m. to 8 p.m. and bottoming in the early morning around 4 – 5 am. Neither of the three types of taxi trips shows very distinct patterns of time of day compared to the overall pattern, except that the transit-extending trips present a noticeable drop after 8 pm, which might be due to travelers' reluctance to take transit late at night though the subway system still operates.

4.3.4 *Spatial Distribution*

The spatial distribution of the three types of taxi trips not only reflects the geographical location of the trips, but also link the trips with built environment factors. Figure 4.8 shows the number of taxi trips' pick-ups/drop-offs at the 200-meter by 200-meter grid level. Transit-complementing taxi trips present the most expansive scale, and the drop-offs of transit-complementing trips are more spread out than the pick-ups. Transit-competing trips follow a similar pattern with subway lines, which is pre-determined by how the trips were classified, but also show a more spread-out pattern of drop-off locations than pick-ups. The drop-offs and pick-ups of transit-extending trips do not vary much, and they concentrate mostly around subway stations. The finding that the drop-offs of taxi trips are more expansive than the pick-ups echoes with previous research that suggested the asymmetrical pattern of taxi trips. The asymmetrical pattern of taxi trips is related to the fact that hailing a taxi is often easier in dense areas concentrated around a smaller number of high-activity centers. Going from high-density to low-density is much easier than the converse. It is also common to get a ride from household member, friend, or transit from the low-density to high-density direction knowing that it will be easy to take a taxi back.

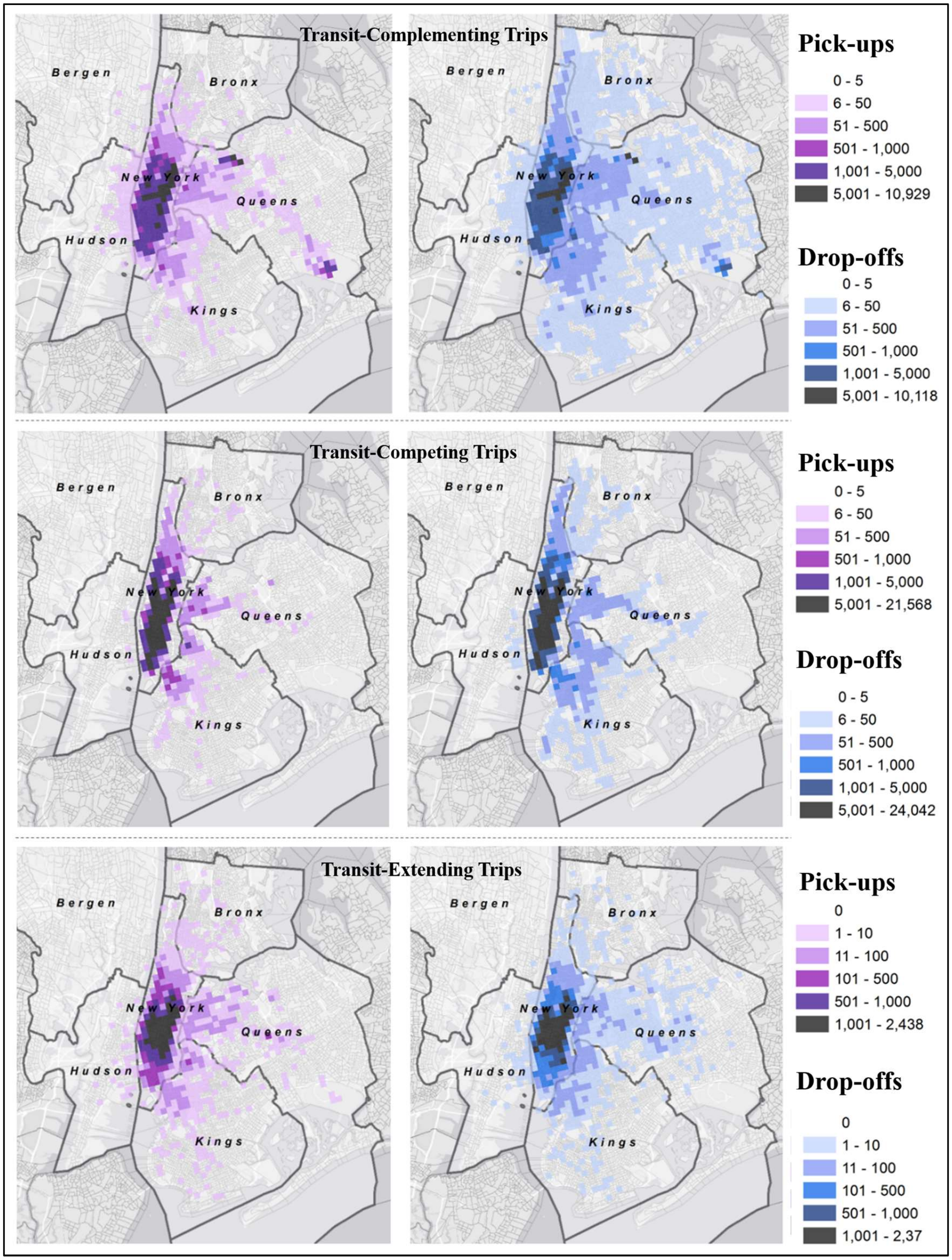


Figure 4.8. Spatial Distribution of the Three Categories of Taxi Trips

4.4 Socio-demographic and Built Environment Characteristics of Places that Generate ODRS Trips

4.4.1 Comparing the Characteristics of Places that Generate Different Types of Taxi Trips

A multinomial logistic (MNL) model is developed to examine what types of places are associated with each of the three classified taxi trips in NYC. The dependent variable is a categorical variable indicating whether the taxi trip is transit-complementing, transit-competing, or transit-extending. Descriptive statistics of the variables included in model are summarized in Table 4.6. The variables included in the model were selected carefully to avoid the issue of collinearity, and different models were selected mainly by comparing their Akaike information criterion (AIC) and log-likelihood values.

There are several interesting findings from examining the results of the MNL model shown in Table 4.7. The reference case of the model is transit-competing trip, so the coefficients need to be interpreted with transit-competing trips as reference. Regarding trip level variables, the three time-related dummy variables, including morning peak, evening peak, and late night are statistically significant. They have same signs for both transit-complementing trip and transit-extending trips. It indicates that transit-competing taxi trips tend to occur less during morning and evening peak hours, and more in late night, while transit-extending and transit-complementing taxi trips take place more often in peak hours and fewer in late night. This indicates that probably during rush hour, trip makers tend to take transit to avoid roadway congestion and thus there is a higher rate of transit-extending taxi trips. Travelers may also not want to take transit at late night, which could be due to longer headways and security concerns. As a result, transit-competing taxi trips happen more in late night, while transit-extending trips decrease significantly at night.

Table 4.6. Descriptive Statistics of Variables Included in the MNL Model

Continuous Variables	mean	sd	min	max
Trip distance (Log-transformed)	2.71	3.08	0.10	82.70
Pick-up: Employment density	35,321	35,696	0	157,639
Pick-up: Employment-population balance	35	174	0	4,181
Pick-up: Median housing value	833,054	213,261	11,700	1,000,001
Pick-up: Poverty rate	0.05	0.09	0	1.00
Pick-up: Bus stop density	8	12	0	98
Pick-up: Bus line density	19,950	13,800	25	41,625
Drop-off: Employment density	35,037	37,090	0	157,639
Drop-off: Employment-population balance	30	177	0	4,181
Drop-off: Median housing value	814,953	224,865	11,700	1,000,001
Drop-off: Poverty rate	0.06	0.10	0	1.00
Drop-off: Bus stop density	7.75	11.25	0	98
Drop-off: Bus line density	25,400	17,375	25	52,975

Categorical Variables	Number of 0s	Number of 1s
Morning peak	845,107	137,975
Evening peak	835,127	147,955
Late night	725,652	257,430
Payment cash	434,766	548,316
Rain dummy	904,125	78,957
Pick-up land use: High residential	817,745	165,337
Pick-up land use: Mixed use	678,605	304,477
Pick-up land use: commercial	677,275	305,807
Drop-off land use: High residential	785,384	197,698
Drop-off land use: Mixed use	716,842	266,240
Drop-off land use: commercial	681,155	301,927

Table 4.7. Results of the MNL Model

	Transit-complementing			Transit-extending		
	Estimate	Std.Error	Pr(>|z|)	Estimate	Std.Error	Pr(>|z|)
(Intercept)	0.325	1.59E-07	***	-1.071	7.79E-08	***
Trip level variables						
Morning peak	0.124	3.08E-08	***	0.189	1.76E-08	***
Evening peak	0.0340	1.38E-08	***	0.0596	1.31E-08	***
Late night	-0.152	7.76E-08	***	-0.221	3.35E-08	***
Trip distance (log-transformed)	0.532	2.06E-07	***	-0.608	5.83E-08	***
Payment cash	0.0300	9.56E-08	***	0.0352	5.04E-08	***
Rain dummy	0.0433	1.43E-08	***	0.0455	6.29E-09	***
Pick-up locational variables						
Pick-up: Employment density	-1.51E-03	2.15E-05	***	-8.88E-04	3.39E-05	***
Pick-up; Employment population balance	2.71E-03	1.28E-04	***	2.40E-03	1.51E-04	***
Pick-up: Poverty rate	-0.486	8.58E-09	***	0.048	4.42E-09	***
Pick-up: Bus stop density	0.328	4.12E-08	***	0.326	2.86E-08	***
Pick-up: Bus line density	9.58E-06	6.84E-07	***	7.92E-06	9.14E-07	***
Pick-up land use: Residential high-density	-0.280	7.28E-08	***	-0.400	2.91E-08	***
Pick-up land use: Mixed use	-0.348	1.38E-07	***	-0.268	6.48E-08	***
Pick-up land use: Commercial	-0.527	2.66E-08	***	-0.169	2.48E-08	***
Drop-off location variables						
Drop-off: Employment density	-1.60E-03	2.18E-05	***	-1.15E-03	3.46E-05	***
Drop-off: Employment population balance	1.69E-03	9.53E-05	***	1.62E-03	1.12E-04	***
Drop-off: Poverty rate	-0.581	1.13E-08	***	0.0218	5.02E-09	***
Drop-off: Bus stop density	0.333	3.84E-08	***	0.362	2.82E-08	***
Drop-off: Bus line density	4.09E-06	4.20E-07	***	3.89E-06	4.65E-07	***
Drop-off land use: Residential high-density	-0.252	6.32E-08	***	-0.322	2.69E-08	***
Drop-off land use: Mixed use	-0.323	8.62E-08	***	-0.267	4.24E-08	***
Drop-off land use: Commercial	-0.488	1.63E-08	***	-0.267	2.52E-08	***
Significance Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
Log-likelihood (null-model) = -109861.2; Log-likelihood (constant-only model) = -87627.27; Log-likelihood (full model) = -75029.98						
Pseudo R2 (null-model based) = 0.317						

The dummy variable indicating whether it was raining (in that hour) when the taxi trip took place is also found to be positively correlated with transit-complementing and transit-extending trips, indicating that when it is raining, there is smaller rate of transit-competing trips. However, this may not reflect the trip maker's willingness to take transit when there is rainfall, but is likely to be a result of the limited supply of taxis during the rain. Kamga, Yazici, and Singhal (2013) identified that drivers make more frequent and slightly shorter trips to increase their income when there is rainfall, and after reaching their income target, drivers may end their shift early, which might be a result of the perceived taxi shortage during prolonged rain conditions. It is likely that the shortage of taxi supply during rainfall resulted in less transit-competing trips. Temperature is not found to have a significant role in the model.

Trip distance is a key variable that distinguishes the three types of taxi trips. The variable of log transformed trip distance has a positive coefficient for transit-complementing trip and has a negative coefficient for transit-extending trip. This is consistent with the trip length distribution shown in Figure 4.7 and reflects that travelers who were farther than two miles from a transit station may not take a taxi to access it (or will not choose to take transit at all). The trip makers' choice of taking a taxi to access transit might be relevant to trip length and the cost of the trip which is nearly proportional to trip length. The data show that about 90% of the transit-extending taxi trips are shorter than 2 miles and about 86.5% of them have a cost below $10.00, indicating that trip makers may prefer taking a taxi to access transit when the trip is shorter than 2 miles with less than a $10.00 cost. Transit-complementing trips tend to have longer trip distances compared to the other two types of taxi trips. Shorter travel distances tend to attract travelers to choose taxi over transit and travelers are more likely to use taxi to access transit for shorter trips, which might be related to cost considerations. Both transit-complementing and transit-extending trips are positively associated with using cash for payment, which may imply that the travelers making these two types of trips comprise more low-income people.

Regarding the built environment factors associated with either the origin or destination of a taxi trip, employment density, median housing value, and poverty rate are found to be statistically significant in the model. This may be related to how the three types of taxi trips are defined, so these locational attributes may be "pre-determined" to some degree. It is consistent with the definition of the three types of taxi trips, as transit-competing trips are more likely to start and end in places with higher density, as those are places with more subway stations. Those places also tend to have mixed land use or commercial land use and it is probably why these variables are identified to be negatively associated with transit-complementing and transit-extending trips in the model. Nevertheless, it is interesting to find the variable of poverty rate has a negative sign for transit-complementing trip but a positive sign for transit-extending trip. This indicates that compared to transit-complementing and transit-competing, transit-extending trips are more likely to start from and end in places with higher poverty rates. The variable of employment-population balance, calculated as employment per person, is found to have positive signs for both transit-complementing and transit-extending trips, which probably implies that better job-housing balance may decrease the probability of people taking a taxi to replace transit.

The findings from the MNL model shown in Table 4.7 suggests that the three types of taxi trips have very different characteristics, which, to some extent, validates the categorization of the trips. It is likely that taxi trips are serving travelers with very different travel needs and are serving areas with very different economic and built environment characteristics. Compared to transit-competing trips, transit-complementing and especially transit-extending trips are more likely to be made by captive users of taxi and may comprise more low-income travelers and travelers residing in peripheral areas.

4.4.2 Comparing Taxi vs. Ride-sourcing Trip Generation

Since the ride-sourcing trip data in NYC including trips made by Uber and Lyft only has information about trips' pick-up location, the classification analysis presented previously cannot be developed for ride-sourcing trips. However, examining the pick-up locations of ride-sourcing trips and comparing the characteristics of places generating ride-sourcing vs. taxi trips can contribute to understanding the difference between ride-sourcing and taxi trips. Therefore, some exploratory analysis and two regression models are developed to reveal whether there is difference in places that generate ride-sourcing trips vs. places that generate taxi trips in NYC.

The descriptive statistics of the number of trip pick-ups made by taxi, Lyft, and Uber aggregated at the block group level are shown in Table 4.8. The numbers of taxi pick-up by block group are significantly larger than trips by Lyft and Uber in 2014-2015. At that time, Uber had been operating for about three years and Lyft had just entered the market in July 2014. This explains why the average number of pick-ups by Uber and Lyft were much smaller compared to pick-ups by taxi and Lyft has the smallest share as shown Table 4.8. There is great variation in the number of ODRS trip pick-ups across different block groups. There are areas that have thousands of shared mobility trips, which mostly concentrate in Manhattan and the JFK airport area, while there are also areas that do not have any ODRS trips. This is probably a combined result of both ODRS supply and demand. On one hand, it is often easier to find ODRS vehicles in those high-density and transportation hub destinations, so travelers can easily use ODRS in those areas. On the other, there are more potential users of ODRS in those areas because of the high-density and high concentration of destinations, so trips starting from those areas are more often found.

Table 4.8. Descriptive Statistics of Block-group Level ODRS Trip Pick-ups

	Average Daily Pick-up or Trip Starts			
	Taxi	Lyft	Uber	Total ODRS (Taxi, Lyft, & Uber)
Mean	53.5	0.6	3.7	57.9
Median	1.1	0.1	0.2	1.4
Standard Deviation	254.4	1.8	18.2	272.6
Minimum	0.0	0.0	0.0	0.0
Maximum	7844.7	40.4	575.8	8119.0

The spatial distributions of taxi pick-ups and ride-sourcing (Uber and Lyft) trip pick-ups in New York City are mapped in Figure 4.9. Although ride-sourcing pick-ups are much fewer than taxi pick-ups overall, ride-sourcing pick-ups have a more expansive spatial distribution, as the trips almost cover all the block groups in the five boroughs of NYC. Many of the peripheral block groups do not have any taxi pick-ups. This may be related to the fact that all ride-sourcing trips are request-based, so a trip request can be met wherever the traveler is, if the region is served by the ride-sourcing company. In contrast, the access to using taxi may be more random, so there may be areas that see fewer or even no taxis idling which make taxi service perceived as unavailable or inaccessible in those areas. Taxi-hailing is easier in dense locations, whereas TNC-hailing is more or less equally easy anywhere in a metro area. For taxi, Uber, and Lyft, lower Manhattan, the JFK Airport area, and areas connecting Manhattan to Brooklyn and Queens have the highest concentration of pick-ups.

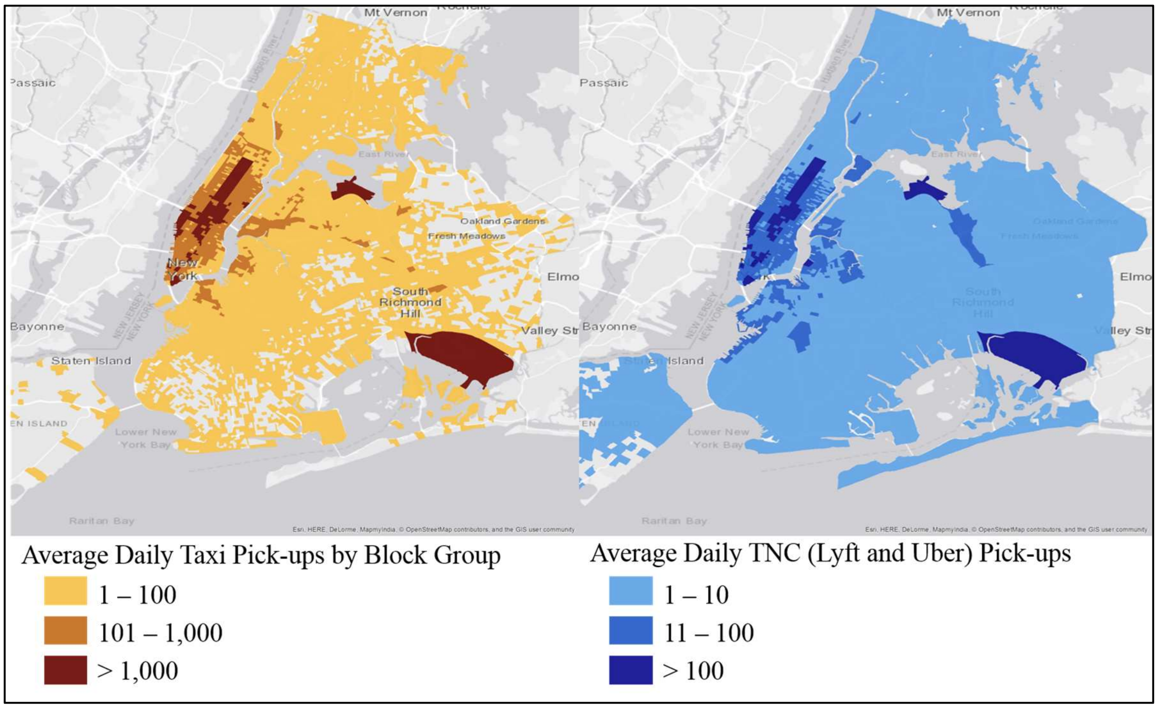


Figure 4.9. Average Daily Taxi Pick-ups and TNC Trip Pick-ups by Block Group

The taxi and ride-sourcing trip data does not include users' profiles, so it does not allow understanding as to what travel demand the trips are serving. However, examining what areas are associated with ODRS trip generation can reveal some underlying dynamics about what travel demand ODRS is serving. Therefore, the block-group level five-year ACS data are used to compare the demographic characteristics of areas that have high vs. low ODRS trips. The results are tabulated in Table 4.9. Areas with higher ODRS trips have significantly higher median household income, a significantly lower African-American population, a significantly higher white population, and higher population density. The vast contrast in the socio-demographic characteristics of places that have high vs. low ODRS trips reveals a potential equity issue.

Table 4.9. Selected Characteristics of Areas with Low vs. High ODRS Trips

	Percent population under 18	Percent population above 65	Median Household Income	African American Population Percentage	White Population Percentage	Population Density (Per Sq. Mile)
Low ODRS Generation (Z-score < 0)	21.20%	12.50%	50,584	26.40%	38.60%	56,310
High ODRS Generation (Z-score >= 0)	12.00%	14.90%	92,643	7.90%	69.20%	100,635

Two simple OLS regression models are developed to examine what built environment and socio-economic factors are associated with taxi trip generation vs. ride-sourcing trip generation and the results are presented in Table 4.10. To compare the two models, the standardized coefficients of the independent variables are presented. As the table shows, the two models for taxi trip generation and ride-sourcing trip generation have very similar results. A same set of independent variables are found to be significant in both models. The first model has an adjusted R-squared value of 0.54 and the second model has an adjusted R-squared value of 0.57, indicating a relatively good proportion of variance explained by the model.

Places with higher population and employment density, better land use mix and job housing balance tend to have higher generation of both taxi trips and ride-sourcing trips. There are relatively more taxi and ride-sourcing trips starting from places with commercial land use and less trip generation from residential areas. For the areas that generate more

ODRS trips, the share of female residents or elderly people is smaller, but the percentage of people with better education attainment is higher. The places with high concentration of ODRS trips tend to have higher median housing price. The standardized coefficients of the independent variables included in both models are very similar. For taxi trip generation, population density and percentage of population with bachelor's degree or higher are the most two important factors. For ride-sourcing trip generation, the percentage of population with bachelor's degree or higher has a much larger standardized coefficient compared to other independent variables and compared to the taxi trip generation model. This is consistent with previous finding that ride-sourcing is serving more well-educated population who are often found to be more tech-savvy and accept new technology faster. This may be also related to the fact that the reliance on using smartphone to access ride-sourcing has created some social, economic, financial and perceptual barriers for equally accessing ride-sourcing service.

The regression models' result shown in Table 4.10 and the descriptive statistics shown in Table 4.9 have suggested that ODRS in New York City is serving more areas that have higher density, higher housing prices, better land use mix, more commercial land development, and higher concentrations of well-educated people, while with less concentration of female and elderly users. Those areas tend to be the central areas where the land and housing values are high and where the level of transit service is high. Contrasting to these characteristics of places with more ODRS trips, there is also evidence that suggests ODRS is serving more transport-disadvantaged population, as described in Section 4.2 and presented in Table 4.3 and Table 4.4. This contrast may imply that there could be some equity issues. On one hand, the ODRS service is provided by private companies seeking profit maximization, so is more likely to concentrate in areas with higher density and higher-income travellers who are likely to be choice users of ODRS. On the other hand, the captive users of ODRS, who tend to be physically, or economically disadvantaged population often live in peripheral areas which may see less supply of ODRS. The supply side of ODRS needs to be researched more to warrant this speculation, but the different travel needs of choice vs. captive ODRS users is an important field for future planning and policy intervention. Currently, most of the ODRS is only provided in urban areas, so how to incentivize ODRS in rural and small urban areas with more transport-disadvantaged people is an important research topic for next steps.

Table 4.10. OLS Regression Result for Taxi Trip vs. Ride-sourcing Trip Generation

	Model 1: Block-Group Level Taxi Pick-ups					**Model 2: Block-Group Level Ride-sourcing Pick-ups**				
	Coeff.	Std. Coeff.	P-value			Coeff.	Std. Coeff.	P-value		
(Intercept)	-1.85		<	0.00	***	-1.81		<	0.00	***
Population density	0.000019	0.38	<	0.00	***	0.000007	0.23	<	0.00	***
Employment density	0.000004	0.12	<	0.00	***	0.000003	0.15	<	0.00	***
% Female residents	-0.64	-0.03		0.00	**	-0.36	-0.03		0.00	**
% Pop older than 65	-2.12	-0.07		0.00	***	-0.96	-0.05		0.00	***
% Pop with bachelor degree or above	4.48	0.40	<	0.00	***	3.69	0.54	<	0.00	***
Land use entropy	2.19	0.22	<	0.00	***	0.87	0.14	<	0.00	***
Job housing balance	1.06	0.12	<	0.00	***	0.62	0.11	<	0.00	***
Land use: residential	-0.98	-0.13	<	0.00	***	-0.57	-0.12	<	0.00	***
Land use: commercial	0.74	0.04		0.00	***	0.36	0.03		0.00	***
	Adjusted R-squared = 0.54					Adjusted R-squared = 0.57				

4.5 Conclusions

The first research question attempts to further the understanding about the role of ODRS in urban transportation by examining most of the publicly available data sources of ODRS trips. First, ODRS trips including both taxi and ride-sourcing trips, are examined to extract the characteristic of ODRS riders and trips. Then a classification analysis is applied to the taxi data in New York City to reveal to what degree taxis are competing with public transportation, versus complementing it or serving the first/last mile of transit. Then regression analysis is developed to identify the characteristics of places that generate different types of taxi trips and the characteristics of places with higher taxi trip generation vs. ride-sourcing trip generation.

The three pieces of analysis reveal three important findings of research question 1. First, the socio-demographic and economic characteristics of taxi riders in New York City and ODRS users nationwide has revealed the role that ODRS has in serving transport-disadvantaged population. It also shows the different market segmentation that ODRS has, including both choice users and captive users who have very different travel needs and socio-economic traits. Second, classifying the taxi trips based on their relationship with transit in New York City reveals that ODRS has different types of impact on the use of transit: about forty percent of the taxi trips are competing with public transportation; about fifty percent are complementing transit; and about seven percent of taxi trips are likely to be made to serve the first/last mile of transit. Focusing on improving the multimodal connection between ODRS and transit across population groups is important to leverage ODRS to improve the benefits that the transport system provides in an equitable way. Third, the regression analysis suggests that there may exist a severe mismatch between

ODRS supply and the potential need for ODRS. The supply of ODRS cater to wealthier places that have higher development density, which lead to more possibility of making ODRS competing with public transportation. However, there is substantial potential need of using ODRS from transport-disadvantaged population who live in areas with less ODRS supply. This is a result of the current unbalanced supply of ODRS which is determined solely by the private sector and implies the necessity of the public sector's intervention.

CHAPTER 5. MODE CHOICE MODELING OF ODRS

Incorporating ODRS into travel mode choice modeling is one of the first steps to incorporate ODRS into normal transportation planning processes. The second research question of the dissertation attempts to use publicly available datasets to explore travel mode choice modeling of ODRS. Travel mode choice is a critical step in travel demand forecasting and the process predicts which travel mode a traveler will choose given a certain set of factors. Using four different datasets, including the 2017 NHTS and three regional household travel survey datasets from the New York region, the Puget Sound region, and the Delaware Valley region, Research Question 2 explores mode choice modeling of ODRS using different statistical and machine learning models.

The mode choice modeling analysis in the dissertation has its limitations. An important limitation is the inability to distinguish between taxi and ride-sourcing trips because of data limitation. Traditional taxis and ride-sourcing are superficially similar with respect to conventionally-measured mode attributes, but the lower wait times and costs alone do not explain the soaring popularity of TNCs over taxi. Convenience, availability, transparency of information, tech-savviness, and "coolness" factors of ride-sourcing are not being accounted for in the models. Another limitation is that the travel mode choices are modeled at the trip level without considering trip chaining effects or tour-level factors. How to incorporate both trip-level and tour-level considerations into travel mode choice analysis has been explored with statistical models, but has not been researched for machine learning models. This may be the next step for facilitating real-world applications of machine learning models to travel mode choice modeling. Another limitation is the small number of observation of ODRS trips in the datasets make the results hard to be generalized. The ability to generalize the results and findings will need to be confirmed by future research when more new data sources become available. Nevertheless, the analysis reveals the great potential of using machine learning for improving travel mode choice prediction accuracy and provides an early contribution to documenting the implementation of the new machine learning model and techniques. The analysis is a starting point to understand ODRS and incorporate it into travel demand forecasting, providing methodological references for future research, and pointing to methodological advancement of travel mode choice modeling.

5.1 Methodology and Data

Research Question 2 of this dissertation intends to identify the factors that relate to people's mode choice of ODRS and update existing travel mode choice models considering the effect of ODRS. Incorporating ODRS into travel mode choice modeling could be the first step of integrating it into travel demand forecasting and everyday transportation planning. In addition to using a multimodal logit (MNL) model, which is one of the most commonly used statistical models for travel mode choice modeling, the dissertation also employs two machine learning models, including an extreme gradient boosting (XGB) model and a random forest (RF) model. The three models are applied to four datasets, including the 2017 NHTS data and the regional household travel survey data from three areas, including the New York metropolitan area, the Puget Sound region, and the Delaware Valley region. Results and performance of using the three models for the four datasets to predict people's travel mode choices considering availability of ODRS are compared. Factors related to people's travel mode choices are identified. The strength and weakness of using statistical models vs. machine learning are discussed.

5.1.1 Data Preparation

The lack of empirical data of ODRS has always been a challenge of conducting research on this topic. Due to the small mode share of ODRS, it is not included in most regions' household travel survey data. The 2017 NHTS data, containing rich information about trips made by taxi and ride-sourcing (Uber and Lyft), provides a great opportunity to study this topic. Thus, this study employs the 2017 NHTS data and three regional household travel survey datasets, including the household survey data from the New York metropolitan area, the Puget Sound region, and the Delaware Valley region, which are the few regions whose household travel survey data contain sufficient number of people using ODRS. The 2017 NHTS data contains more than two thousand observations of trips made by taxi and ride-sourcing, which provides a great opportunity to research ODRS-related inquiries. The NYMTC RHTS data were collected during 2010 and 2011 in a 28-county area of New York, New Jersey, and Connecticut (hereafter as "New York region"). The dataset contains 143,925 linked trips from 18,965 households and 43,558 participants from the metropolitan area shown as Figure 5.1. Taxi and for-hire transportation service were included in this dataset. The 2014-2015 Puget Sound Regional Travel Survey (PSRTS) data were collected in 2014 from the Puget Sound region (hereafter as "Puget Sound region"), as shown Figure 5.2. It is one of the most recent travel survey data of metropolitan areas that include specific identification of both taxi and ride-sourcing as travel modes. The Delaware Valley Regional Planning Commission (DVRPC) household travel survey data were collected in the Delaware Valley region in 2012. The Delaware Valley region consists of nine counties in Pennsylvania and New Jersey as shown in Figure 5.3. The dataset contains 81,940 unique trip records, but more than 20,000 of them omit the information on travel mode and about 6,000 omit the information of income levels. After the data cleaning process that removes the trips with missing values in the dependent and independent variables we want to model, there are 51,910 trips made by car, biking, walking, or transit that could be used for the mode choice analysis.

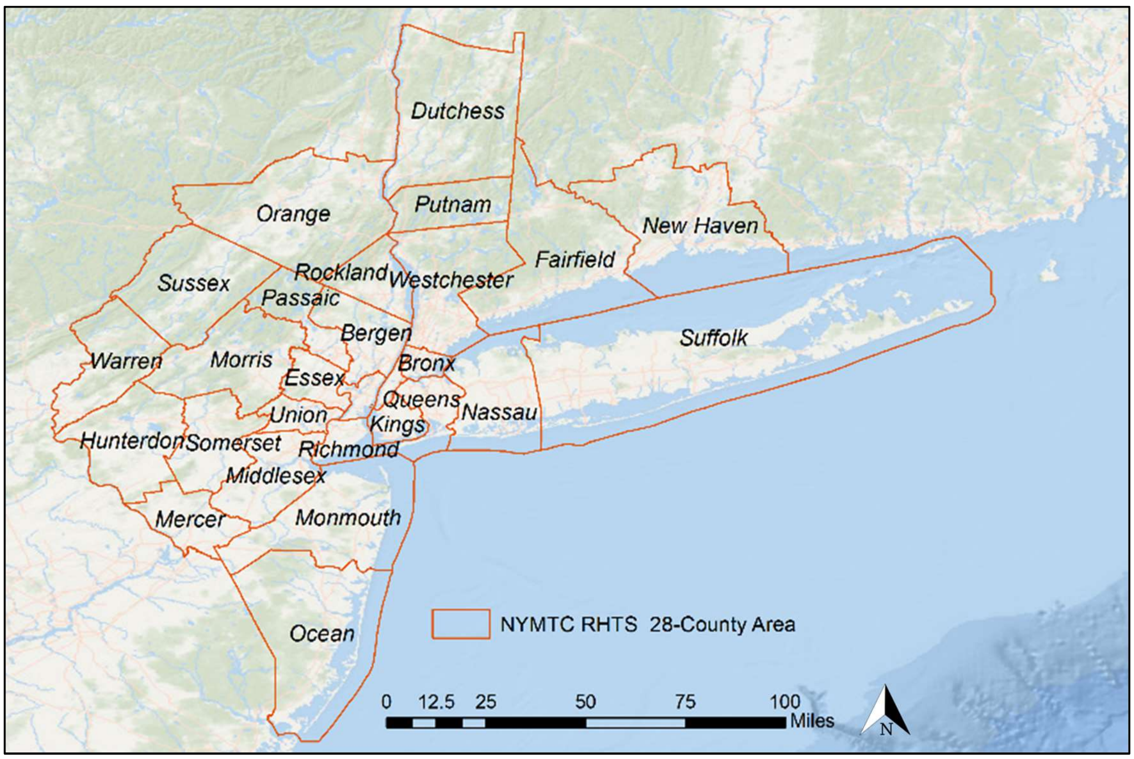


Figure 5.1. The 28-County Area of the New York Metropolitan Area

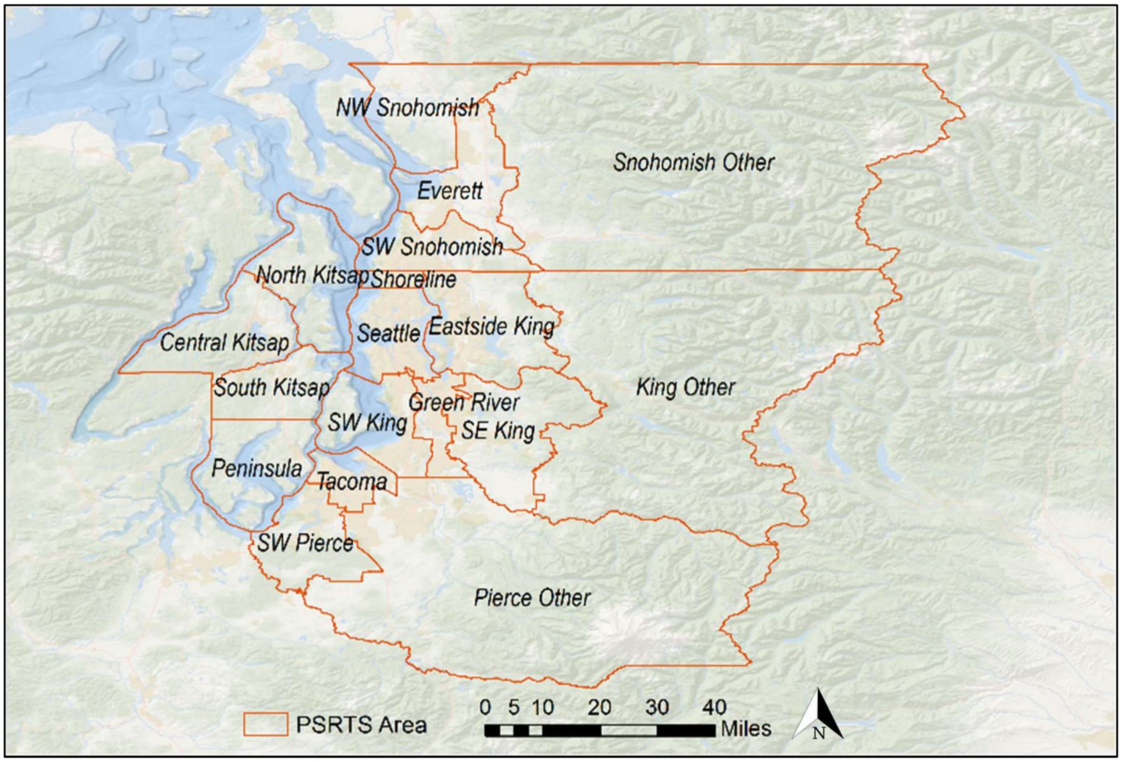


Figure 5.2. The Puget Sound Region

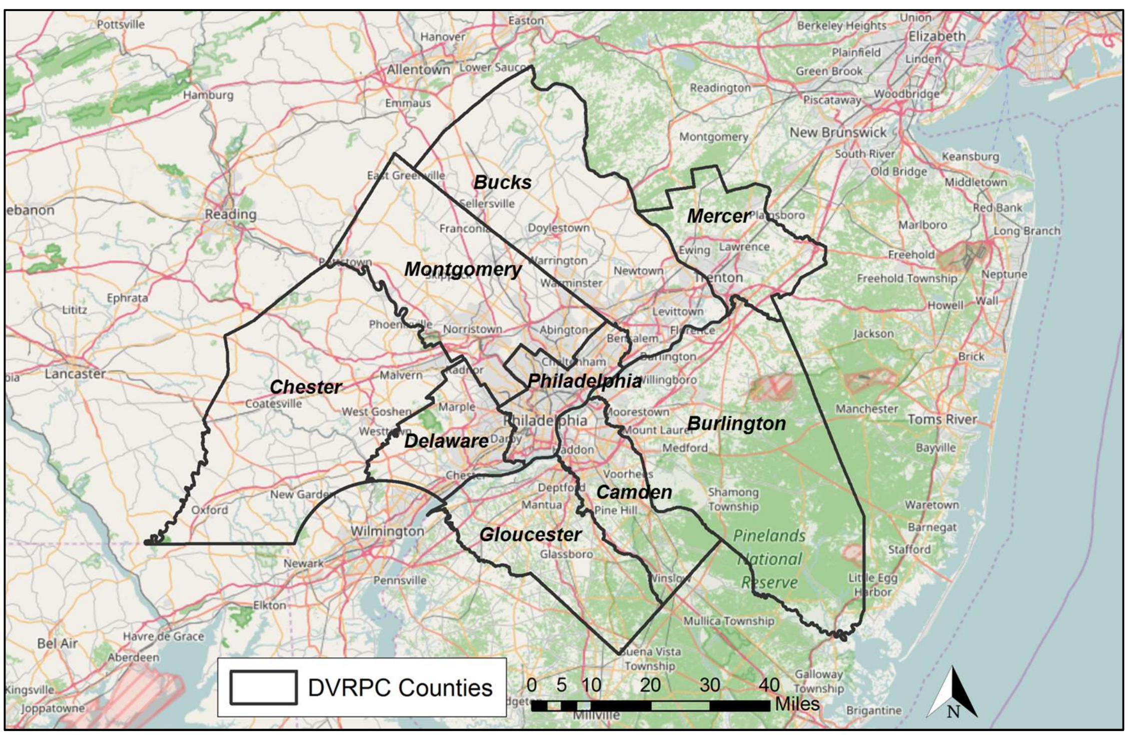


Figure 5.3. The Delaware Valley Region

Table 5.1 summarizes the independent variables included in the mode choice models developed using the four different datasets, their explanation, and data sources. For all the models, the dependent variable is people's mode choice indicating whether the trip is made by car (driving or passenger), biking, walking, fixed-route transit, or ODRS. The independent variables include three types of factors suggested by existing literature, including trip characteristics, person/household features, and neighborhood-level variables. Most of the variables are calculated using publicly available datasets, as shown in the table, except that the travel time of alternative transportation modes are requested using the Google Maps Distance Matrix API. The regional household travel survey datasets of the two regions have the variables of reported travel time, but they only has the travel time for the "chosen" mode. However, to implement discrete choice models, it is also necessary to know the travel time of the "unchosen" modes to form travelers' complete choice sets. The Google Maps Distance Matrix API provides estimated travel time by driving, biking, walking, and transit at different times of day. Thus, the departure time of each trip and the trip origin/destination (centroids of census tracts for the New York region, centroids of block groups for the Puget Sound region, and centroids of TAZs for the Delaware Valley region) were used to request the estimated travel time of all the "unchosen" cases. The travel time by ODRS is computed as the travel time of driving plus a random waiting time ranging from 1 minute to 10 minutes, since Rayle et al., (2016) found that more than 90% ride-sourcing trips have wait times shorter than 10 minutes.

The 2017 NHTS data does not have information about trips' origin or destination, so Google Maps API cannot be used to request corresponding travel time in the national model. The travel time of the unchosen mode in the national model is thus estimated mainly according to some simple rules (e.g. whether there is public transit), travel distance, and the average travel speed of different modes. Since the 2017 NHTS data does not have information about trips' origin or destination, most of the built environment variables cannot be computed, and therefore only population density is included as a neighborhood variable in the national model.

Table 5.1. Variables Included in Mode Choice Models

Variable Name	Explanation	Data Source
Travel mode	Travel mode (dependent variable): 1= car; 2 = walk; 3 = bike; 4 = fixed route transit; 5 = ODRS	Four travel survey datasets
Trip Variables		
Travel time	Travel time by mode in minutes	Four travel survey dataset and Google Maps Distance Matrix API data
Trip distance	Trip distance in miles	Four travel survey datasets
Trip cost	Trip cost by mode in dollar	
Morning peak	The trip is happening in the morning peak hour (7am - 9am)	Four travel survey datasets. The appearance of a subscript "(1) (2) (3) or (4)" indicates that this variable is not included in all three models. "(1)" indicates this variable is included in the national model; "(2)" indicates that the variable is included in the New York model; "(3)" indicates that the variable is included in the Puget Sound model; and "(4)" indicates that the variable is included in the Delaware Valley model;
Evening peak	The trip is happening in the evening peak hour (4pm – 7pm)	
Late night	The trip is happening in late night (9pm - 4am)	
Activity duration	Activity duration in minutes	
Trip purpose home (2)(3)	The trip purpose is home	
Trip purpose work (2)(3)	The trip purpose is work	
Trip purpose recreation (2)(3)	The trip purpose is recreation	
Trip purpose maintenance (2)(3)	The trip purpose is maintenance	
Trip purpose school (2)(3)	The trip purpose is school	
Trip purpose change mode (2)(3)	The trip purpose is to change a travel mode	
Home-based work (4)	The trip is home-based trip made for work	
Home-based other (4)	The trip is home-based for other purpose	
Total travelers	Total number of travelers in the trip	

(Table 5.1 Continued)

Person/Household Variables		
License	The traveler has a driver’s license	Four travel survey datasets
Low income	The traveler is low income (annual income < $30,000)	
High income	The traveler is high income (annual income >= $100,000)	
Young	The traveler is younger than 16	
Elderly	The traveler is elderly than 65	
Disability (2)(4)	The traveler has disability	
Female	The traveler is female	
Household size	The traveler's household size	
Employed	The traveler is employed	
Student	The traveler is a student	
Number of vehicles per capita	Number of vehicles per capita in the household	
Education attainment (4)	Education attainment level of the traveler	
Park must pay (4)	Traveler must pay for parking	
Employer subsidize transit (4)	Employer provides subsidy for transit	
Life cycle	Life cycle variables indicating whether the traveler is married and/or have children	
Neighborhood Variables (Calculated at census tract level for the New York and Delaware Valley regions and at the block group level for the Puget Sound region)		
O/D: Population density	Population density (persons per sqml) at the origin or destination	ACS 5-year estimate 2011-2015 (2017 NHTS data for the national model)
O/D: employment density (2)(3)(4)	Employment density (jobs per sqml) at the origin or destination	LEHD 2011 data
O/D: employment entropy (2)(3)(4)	Employment diversity (3-category entropy*) at the origin or destination	LEHD 2011 data
O/D: job-housing balance (2)(3)(4)	Employment population balance (entropy metric) at the origin or destination	Calculated using the ACS and LEHD data above
O/D: bus stop density (2)(3)(4)	Bus stop density (# bus stops per sqkm) at the origin or destination	Calculated using GTFS data
O/D: subway stop density (2)(3)(4)	Subway stop density (# subway density per sqkm) at the origin or destination	Calculated using GTFS data
O/D: road density (2)(3)(4)	Road density density (road length in km per sqkm) at the origin or destination	Calculated using OpenStreetMap data - Roads
Land use diversity (2)(3)(4)	Entropy index of land use diversity	Calculated using land use data
Median housing value (2)(3)(4)	Median housing value	ACS 5-year estimate 2011-2015

*Employment entropy is calculated as an entropy index considering three types of employment: retail, service, and finance.

5.1.2 Data Sampling

A major challenge of modelling the travel mode choice of ODRS is its extremely small mode share. The extremely small share of ODRS in household travel survey data results in an unbalanced data issue which will often result in poor model fitting. For example, when different classes are represented very unequally, the estimation of the MNL model may be biased resulting in particularly higher prediction error for classes with smaller shares. Five travel modes are considered in the analysis, including car (driving or passenger), biking, walking, transit, and ODRS. The total number of unique trips included in the analysis after removing missing and outlier values and the mode shares of the five modes are shown in Table 5.2. As shown in the table, the shares of ODRS in four datasets are extremely small, making the datasets very unbalanced. The mode share of ODRS is about 0.3% in the NHTS data, about 0.8% in the New York metropolitan area, and about 0.2% in both the Puget Sound and 0.4% in the Delaware Valley Region. Bike is another travel mode with very small mode shares. Trips made by biking account for about 0.9% in the national data, 0.5% in the New York region, 2.0% in the Puget Sound region, and 1.0% in the Delaware Valley region.

Table 5.2. Mode Shares in the Regional Household Travel Survey Data

		Total Unique Trips	By Car	By Biking	By Walking	By Transit	By ODRS*
National	Number	876,746	773,770	7,872	79,284	13,070	2,750
	Share	100.0%	88.3%	0.9%	9.0%	1.5%	0.3%
New York Region	Number	167,780	97,147	783	47,108	21,408	1,334
	Share	100.0%	57.9%	0.5%	28.1%	12.8%	0.8%
Puget Sound Region	Number	46,036	33,052	921	8,747	3,222	94
	Share	100.0%	71.8%	2.0%	19.0%	7.0%	0.2%
Delaware Valley Region	Number	52,100	43,196	511	5,467	2,736	190
	Share	100.0%	82.9%	1.0%	10.5%	5.3%	0.4%

* The national data and the Puget Sound data include trips made by both taxi and ride-sourcing, while the New York and Delaware Valley data only have taxi trips

The dissertation starts with developing the mode choice models using the four original datasets whose mode shares are presented in Table 5.2 and this showed a serious issue of high predicting error for mode choice of ODRS and biking. In general, the predicting errors for the choice of car, walking and transit could achieve 20% or lower, but the predicting errors for choice of ODRS and biking are as high as 90%. Therefore, rather than using the original regional household travel survey datasets, subsamples of the datasets that contain more evenly distributed mode shares are used in the travel mode choice modeling. The original datasets are sampled following two steps: (1) all the trips made by ODRS are retained; (2) if the trips made by certain mode are less than ODRS trips, all the trips of that mode are also retained; (3) for the rest of the travel modes, the original trips are randomly sampled to make the subsamples account for about 10% of the original datasets. Sampling is a commonly used technique in dealing with unbalanced data issue. The unbalanced data issue can also be dealt with adding some weight parameters to the model or using weight balancing techniques in machine learning. Considering the objective of comparing the models' performance, it is more straightforward to sample the data rather than adding extra parameters to the models that may create inconsistency between statistical modeling and machine learning. The sample sizes and mode shares after sampling the original datasets are shown in Table 5.3. After sampling, trips made by ODRS account for 14.9%, 8.3%, 2.0%, and 3.8% out of all trips in the four subsampled datasets.

Table 5.3. Mode Shares after Resampling

		Total Unique Trips Included in the Model	By Car	By Biking	By Walking	By Transit	By ODRS*
National	Sample Size	18,457	3,986	3,979	3,958	3,784	2,750
	Share	100.0%	21.6%	21.6%	21.4%	20.5%	14.9%
New York Region	Sample Size	16,000	3,965	783	8,137	1,781	1,334
	Share	100.0%	24.8%	4.9%	50.9%	11.1%	8.3%
Puget Sound Region	Sample Size	4,500	3,086	200	799	325	90
	Share	100.0%	68.6%	4.4%	17.8%	7.2%	2.0%
Delaware Valley Region	Sample Size	5,000	2,299	511	1,000	1,000	190
	Share	100.0%	46.0%	10.2%	20.0%	20.0%	3.8%

* The national data and the Puget Sound data include trips made by both taxi and ride-sourcing, while the New York and Delaware Valley data only include taxi trips

5.1.3 Models and Implementation

Travel mode choices have been widely modeled with MNL models in both academia and practice. The well-founded theory of applying MNL model to discrete choice analysis and its desirable closed form of mathematical estimation are the main merits of

MNL models. The main constraint of developing MNL models is its high demand for data quality and its stringent statistical assumptions that often require a careful model specification process. Machine learning models have shown their merits in more flexibility in dealing with nonlinear data relationships and less effort required for variable selection. The major drawback of machine learning models are their poor explanatory powers that often do not allow quantitative interpretation of model results. The MNL model and two machine learning models, including a Random Forest (RF) model and an Extreme Gradient Boosting (XGB) model, are applied to travel mode choice modeling using the regional household travel survey data from the New York metropolitan region, the Puget Sound region, and the Delaware Valle region. The results of the three models are compared regarding their predictive power and their interpretation implications. Both the RF model and the XGB model are tree-based ensemble models and they are used in this analysis mainly because that tree-based models allow understanding the importance of an independent variable in influencing the dependent variable. Some other popular machine learning models, such as the neural network model is operating like a black box and is completely unable to be interpreted, so the RF and XGB models are preferred here.

5.1.3.1 Multinomial Logit Model

The MNL model is the mostly widely used model structure for travel mode choice modeling. It is based on the random utility theory that assumes the utility of choosing a certain travel mode is a random variable that travelers always want to maximize. The utility of choosing a travel mode i can be denoted as Equation (4.1) as follows.

$$U_i = V_i + \varepsilon_i \qquad \text{Equation (5.1)}$$

Where U_i is the utility of choosing a certain travel mode that can be expressed as some random variables V_i and the error term ε_i

The probability of choosing the ith mode from a set of n travel alternatives is thus:

$$P_i = Pr[U_i > U_j] = P_r[\varepsilon_i < V_i - V_j + \varepsilon_j]\ (j \neq i), \qquad \text{Equation (5.2)}$$

where j is another travel mode from the alternative set.

Equation (5.1) and (5.2) form the foundation of random utility theory and based on different assumptions of distribution of the error terms in Equation (5.1), there form different models. An MNL model assumes that the error terms are independent and identically distributed (iid) and they follow a Gumbel distribution. It also assumes that there is homogeneity in responsiveness to attributes of alternative across individuals and the error variance-covariance structure of the alternatives is identical across individuals. These three assumptions lead to the simple and elegant closed-form mathematical structure of the MNL (Bhat, 2003) and the estimation of the MNL model allows direct interpretation, both of which are the reasons for its common application in academia and travel demand forecasting practice. One major constraint of the MNL model is that it assumes the "independence of irrelevant alternatives" (IIA) property (Ben-Akiva & Lerman, 1985; Bhat, 2003).

The multinomial logit model is implemented with maximum likelihood estimation using the "mnlogit" package in R. The specification of an MNL model is often critical for its holding of assumptions. The "mnlogit" package allows the specification of three types of variables, the generic variables, the individual-specific variables, and the alternative specific variables. In this study, except that the variables of travel time and cost (by mode) are either set as alternative-specific or generic variables, all other variables are included in the model as individual-specific variables as their values only vary by individuals and not by travel alternatives. Variables selection was performed based on three criteria: (1) a variable's sign needs to be consistent with existing theory; (2) the variable is statistically significant; and (3) groups of variables are selected by developing chi-squared tests to compare whether adding a group of variables improves the goodness of fit of the model. The model prediction was implemented by conducting a Monte Carlo simulation process: first, applying the model's estimation result back to the data gives a predicted probability for each travel mode of a trip (the predicted probabilities of all modes of the same trip adds up to one); then a random number between zero and one is assigned to the trip and which mode's cumulative probability range this random number falls into, the trip will be forecasted as using that mode.

Since different trips in the dataset might be associated with the same person, which may violate the assumption that the variables are independent and identically distributed in the MNL model. Therefore, in addition to developing standard MNL models, the models are also developed using cluster-robust standard errors to compensate for violation of the independence of observations. Models with cluster-robust standard errors are developed using the "clusterSEs" package in R.

5.1.3.2 Random Forest Model

The Random Forest (RF) model is widely used in machine learning for regression and classification in recent years and it is an ensemble method based on the decision tree algorithm. A decision tree is a non-parametric supervised learning method that is based on a tree-like format of predicting the dependent variable according to simple decision rules of different independent variables. The decision tree model is often considered as more unbiased when it is deeper (more levels of leaves and nodes), but in practice, the decision tree often has large variance and is sensitive to small changes in the data. An RF model is built on the prediction of a set of decision trees and for classification problems, the final prediction of an RF model depends on the voting of all the predictions of the decision trees. By collectively "averaging" the results of a set of decision trees, an RF model can reduce the variance compared to a decision tree model and is more robust to data changes. One advantage of tree-based classification models is that the result can be interpreted to some degree. A decision tree model allows direct interpretation of its result and an RF model, though cannot be interpreted directly, allows the understanding of what independent variables play important roles in predicting the dependent variable.

The RF model is implemented using the "randomForest" package in R. Parameters are tuned (optimized) using cross-validation in every run of the RF model, including the parameter 'mtry' that indicates the number of variables randomly sampled as candidates at each split and the parameter of 'nodesize' that sets the minimum size of terminal nodes. The parameter 'ntree' indicating the number of trees generated in a RF model is fixed as 1000, in every run.

5.1.3.3 Extreme Gradient Boost Model

The Extreme Gradient Boost (XGB) Model is also a tree-based ensemble method that can be used for both regression and classification. The XGB model was first proposed by (Friedman, 2001). It is an ensemble method that is also built upon decision tree algorithm, which is like the RF model. However, the XGB model differs from the RF model in the way that the decision trees are developed. For an RF model, the trees are developed in parallel, but for an XGB model, the trees are developed iteratively based on the idea of additive training. To put it simply, in an XGB model, each decision tree is built to minimize a defined loss function. Each time the estimation puts more weights on the cases that are wrongly predicted by previously developed trees. The final model result is collectively determined by the results of all the developed trees. Like the RF model, the result of an XGB model, allows the understanding about what factors are important for predicting the dependent variable, though they cannot be quantified for interpretation.

The XGB model is implemented using the "XGboost" package in R. Compared to the RF model that is easily tuned, the XGB model has more parameters that need to be tuned for each run. The parameters that are tuned include:

1) "nrounds": the maximum number of iterations (like the number of trees to grow in our case);
2) "eta": it controls the learning rate;
3) "max_depth": it controls the depth of the tree;
4) "min_child_weight": it controls the minimum value of sum of instance weight of a node;
5) "subsample": it controls the number of samples supplied to a tree;
6) "colsample_bytree": it controls the number of variables supplied to a tree;
7) "max_delta_step": it controls regularization and is used to avoid overfitting.

To optimize the performance of the XGB model and to avoid the overfitting problem, the parameters of “nrounds” and “eta” were tuned first. The tuning result suggested that setting “nrounds” as 100 for the New York model and 50 for the other three models should be effective to avoid overfitting. The parameter of “eta” is set to 0.2 for both models. Then cross-validation is used to tune all the other parameters except “max_delta_step” for every run. The “max_delta_step” was set as 1 for both models as it improves the performance of predicting rare cases (the travel modes with small shares). The parameter of “eval_metric” that defines the loss function is set as the multiclass classification error rate calculated as number of wrong cases divided by the number of all cases.

5.1.3.4 Comparing Model Performance

Each of the three models are run 100 times using each of the four datasets to compare their average prediction accuracy and robustness to data changes. For each run, the dataset was randomly split into a training subset (75% of the data) and a testing subset (25% of the data) and the training errors and testing errors are averaged for the 100 times run. To compare the predictive power of the MNL model and the two machine learning models, the average total errors of the models are recorded, and travel-mode-specific errors are also recorded. The total error is calculated as the number of trips that are predicted to have the wrong mode choice out of the total number of trips. The mode-specific prediction error is calculated as the number of trips wrongly predicted out of the total number of trips made by that mode.

5.2 Descriptive Analysis

As illustrated more in-depth in the Methodology chapter (Section 5.1), the dependent variable of the discrete choice models is the choice among five travel modes: car, biking, walking, transit, and ODRS. The four household travel survey datasets are sampled to the make the shares of different travel modes more evenly distributed. Three groups of independent variables including the trip variables, personal-household variables, and neighborhood variables are included in the analysis, as they are the main factors that are associated with people's travel mode choice as suggested by the literature. Descriptive statistics of the independent variables included in models using the four different datasets are presented respectively in Table 5.4, Table 5.5, Table 5.6, and Table 5.7. Some independent variables are only included in one or two of the models depending on different data availability for different regions. As the tables show, there is significant variation in the values of most variables. This is because the four datasets cover large geographical regions, where the development patterns, socio-demographic characteristics, and travel behaviors are very diverse.

Table 5.4. Variables in the Model Using the 2017 NHTS Data

N	Trips by Car	Trips by Biking	Trips by Walking	Trips by Transit	Trips by ODRS
18,457 (100%)	3,986 (21.6%)	3,979 (21.6%)	3,958 (21.4%)	3,784 (20.5%)	2,750 (14.9%)
Continuous Variables					
		mean	**s.d.**	**min**	**max**
Travel time (minutes)		37.67	79.82	0.03	987.51
Trip distance (miles)		6.07	13.75	0	323.96
Trip cost (dollar)		6.33	20.72	0	813.27
Number of travelers		1.87	3.76	1	221
Household size		2.48	1.37	1	10
Household vehicles per driver		0.96	0.87	0	20
Age		45.52	19.98	5	92
Categorical Variables					
			Count	Share	
Trip Departure Time		Morning peak	2768	15%	
		Evening peak	3691	20%	
		Late night	1107	6%	
Loop trip			1107	6%	
Trip Purpose		Home	7013	38%	
		worker	2399	13%	
		School	738	4%	
		Medical	369	2%	
		Shopping	2399	13%	
		Social	2953	16%	
		Meals	1107	6%	
Weekend			3691	20%	
Low-income			3875	21%	
High-income			6644	36%	
Lifecycle		No children in the household	12550	68%	
		Household has one adult	5352	29%	
		Household has two adults and children	4983	27%	
Socio-Demographic		Low population density at the trip maker's home location	4245	23%	
		High population density at the trip maker's home location	8859	48%	
		Education attainment: high school or below	3875	21%	
		Education attainment: bachelor's degree or above	9413	51%	
		Female	8859	48%	
		Worker	10335	56%	
		Elderly	3691	20%	
		Retired	4429	24%	
		Flexible work time	5721	31%	
		Use smartphone everyday	14138	77%	
		Medical device used	1476	8%	

Note: "O" and "D" stand for trip "origin" and "destination"

Table 5.5. Variables in the Model Using the 2011 RHTS Data

N	Trips by Car	Trips by Biking	Trips by Walking	Trips by Transit	Trips by ODRS
16,000	3,965	783	8,137	1,781	1,334
100.00%	24.80%	4.90%	50.90%	11.10%	8.30%
Continuous Variables					
		mean	**s.d.**	**min**	**max**
Travel time (minutes)		18	24	0	825
Trip distance (miles)		4	7	0	92
Trip cost (dollar)		2	5	0	149
Activity duration (minutes)		201	249	1	1,349
Number of travelers		2	1	1	81
Household size		3	1	1	10
No. of children in the household		1	1	0	8
Vehicle ownership per capita		1	1	0	8
O/D: population density		28,559	37,995	0	241,752
O/D: employment density		64,422	191,022	0	3,153,309
O/D: employment entropy		1	0	0	1
O/D: employment population balance		1	0	0	1
O/D: bus stop density		76	308	0	6,079
O/D: subway/rail station density		4	11	0	74
O/D: roads density		42,870	21,490	2,478	225,752
O/D: median housing value		505,400	354,031	0	2,000,001
Categorical Variables					
			Count	**Share**	
Trip Departure Time		Morning peak	4,160	26%	
		Evening peak	4,640	29%	
		Late night	960	6%	
Trip Purpose		Home	4,640	29%	
		Work	2,080	13%	
		Recreation	2,240	14%	
		Maintenance	2,080	13%	
		Change travel mode	3,520	22%	
Social Demographic		Low income	3,040	19%	
		High income	6,560	41%	
		Young	2,080	13%	
		Elderly	1,600	10%	
		Disability	960	6%	
		Female	8,480	53%	
		Employed	9,920	62%	
		Student	1,920	12%	

Note: “O” and “D” stand for trip “origin” and “destination”

Table 5.6. Variables in the Model Using the 2014 Puget Sound Data

N	Trips by Car	Trips by Biking	Trips by Walking	Trips by Transit	Trips by ODRS
4,500	3,086	200	799	325	90
100.00%	68.60%	4.40%	17.80%	7.20%	2.00%
Continuous Variables					
		mean	**s.d.**	**min**	**max**
Travel time (minutes)		95	202	0	1,427
Trip distance (miles)		5	7	0	71
Trip cost (dollar)		4	9	0	195
Activity duration (minutes)		218	243	1	1,355
Number of travelers		2	1	1	9
Household size		2	1	1	8
Vehicle ownership per capita		1	0	0	5
O/D: population density		8,763	10,124	0	141,622
O/D: employment density		18,567	64,629	0	722,004
O/D: employment entropy		1	0	0	1
O/D: employment population balance		1	0	0	1
O/D: bus stop density		105	158	0	1,017
O/D: subway/rail station density		0	1	0	10
O/D: roads density		55,427	31,178	3,182	218,346
O/D: median housing value		348,299	198,500	0	1,881,600
Categorical Variables					
		Variable	**Count**	**Share**	
Trip departure time		Morning peak	810	18%	
		Evening peak	1,395	31%	
		Late night	225	5%	
Trip purpose		Home	1,620	36%	
		Work	720	16%	
		Recreation	990	22%	
		Maintenance	675	15%	
		Change mode	0	0%	
Income level		Low	450	10%	
		High	1,755	39%	
Age		Younger than 18	450	10%	
		Elderly than 65	675	15%	
Female			2,385	53%	
Employed			2,880	64%	
Student			315	7%	
Have a driver's license			3,825	85%	

Note: "O" and "D" stand for trip "origin" and "destination"

Table 5.7. Variables in the Model Using the 2011 Delaware Valley Data

N	Trips by Car	Trips by Biking	Trips by Walking	Trips by Transit	Trips by ODRS
5,000	2,299	511	1,000	1,000	190
100.00%	46.00%	10.20%	20.00%	20.00%	3.80%
Continuous Variables					
		mean	**s.d.**	**min**	**max**
Travel time (minutes)		14	16	0	220
Trip distance (miles)		5	8	0	76
Trip cost (dollar)		2	4	0	73
Activity duration (minutes)		125	172	1	960
Number of travelers		1	1	1	7
Household size		3	1	1	10
Vehicle ownership per capita		1	0	0	8
O/D: population density		46,315	94,383	0	506,607
O/D: employment density		379,084	1,544,565	11	9,515,692
O/D: employment entropy		1	0	0	1
O/D: employment population balance		1	0	0	1
O/D: bus stop density		435	1,053	0	4,790
O/D: subway/rail station density		30	103	0	501
O/D: land use entropy index		0.66	0.23	0	1
Categorical Variables					
			Count	**Share**	
Trip departure time		Morning peak	1,150	23%	
		Evening peak	1,250	25%	
		Late night	250	5%	
Trip type (purpose)		Home-based work	2,050	41%	
		Home-based other	2,600	52%	
Income level		Low	900	18%	
		High	1,800	36%	
Age		Younger than 18	550	11%	
		Elderly than 65	1,000	20%	
Disability			200	4%	
Female			2,650	53%	
Student			950	19%	
License			3,950	79%	
Must pay for parking			850	17%	
Employer transit subsidy			400	8%	
Education: bachelor's degree or above			2,900	58%	

Note: "O" and "D" stand for trip "origin" and "destination"

5.3 Models' Prediction Performance and Results

For each of the four datasets, including the 2017 NHTS data, and the regional household travel survey data from the New York region, the Puget Sound region, and the Delaware Valley region, each of the three models, including the MNL model, XGB model, and RF model, were run 100 times to compare the models' performance and robustness. For each run, the dataset was randomly split into a training subset (75% of the whole data) and a testing subset (25% of the whole data) and the training errors and testing errors are averaged for the 100 times run. To thoroughly compare the models' performance, not only the overall errors are computed, the mode-specific errors are also computed. As shown in the tables, both the average training and testing errors are computed for all five modes combined and for each of the five modes. The variance of the errors is recorded to examine whether the models' performance is robust to data changes.

The average training and testing errors of the three models using the 2017 NHTS data are presented in Table 5.8 and Figure 5.4 presents a quick comparison of the models' prediction errors. Several findings can be derived by comparing the performance of different models and mode-specific errors. First, the two machine learning models have much lower prediction errors compared to the MNL model and the RF model has the best performance among the three models. The MNL model has an overall prediction (testing) error of about 57.5%, the XGB model has an overall error of 42.9%, and the RF model's overall error is only 23.1%. This indicates that while the MNL model can only correctly predict nearly 42% of the mode choice of all trips, the RF model is able to correctly predict 77%. Second, all the three models perform worst in predicting the choice of ODRS, as ODRS-specific testing error is 71.7%, 68.4% and 38.6% for the MNL, XGB, and RF model respectively. The models also perform poorly in predicting the choice of biking. The higher prediction error of the choice of biking and ODRS may be a result of the smaller shares of the two modes and could also be because that why people choose these two modes are not captured well in these models. The RF model also has significantly lower biking- and ODRS- specific prediction errors. Third, the mode-specific prediction errors of the models are not necessarily consistent with the magnitudes of the mode's share. It means that though the choice of car and the choice of walking account for similar proportions in the data, the car-specific error is much higher than the walking-specific error. This indicates that the models are better at explaining the choice of certain modes than others, which could be a result of variable selection and data availability or could be related to the heterogeneity in the choice of certain modes. Fourth, the RF model and the MNL model are more consistent in training and testing errors indicating they may have better generalizability. One commonly known drawback of machine learning model is the difficulty of avoiding the overfitting problem. The RF model is very good at avoiding the overfitting problem by design, but overfitting can be an issue for training the XGB model. Finally, the variance in the errors of the three models are very small, indicating that all three models are very robust to data changes.

Table 5.8. Models' Performance for the National Data

2017 NHTS		MNL		XGB		RF	
		Training Error	Testing Error	Training Error	Testing Error	Training Error	Testing Error
Total	Mean	57.4%	57.5%	34.3%	42.9%	23.1%	23.1%
	Variance	0.0000	0.0001	0.0024	0.0007	0.0132	0.0126
Car	Mean	58.5%	58.7%	28.2%	36.2%	19.5%	19.4%
	Variance	0.0001	0.0003	0.0018	0.0007	0.0076	0.0069
Bike	Mean	61.6%	61.6%	45.7%	55.0%	26.0%	26.0%
	Variance	0.0001	0.0004	0.0056	0.0021	0.0248	0.0247
Walk	Mean	40.8%	41.1%	16.0%	22.8%	11.9%	12.1%
	Variance	0.0001	0.0003	0.0006	0.0002	0.0026	0.0029
Transit	Mean	58.7%	58.5%	28.2%	40.3%	24.6%	25.1%
	Variance	0.0001	0.0003	0.0018	0.0004	0.0122	0.0127
ODRS	Mean	71.8%	71.7%	55.6%	68.4%	37.7%	38.6%
	Variance	0.0001	0.0003	0.0049	0.0010	0.0359	0.0348

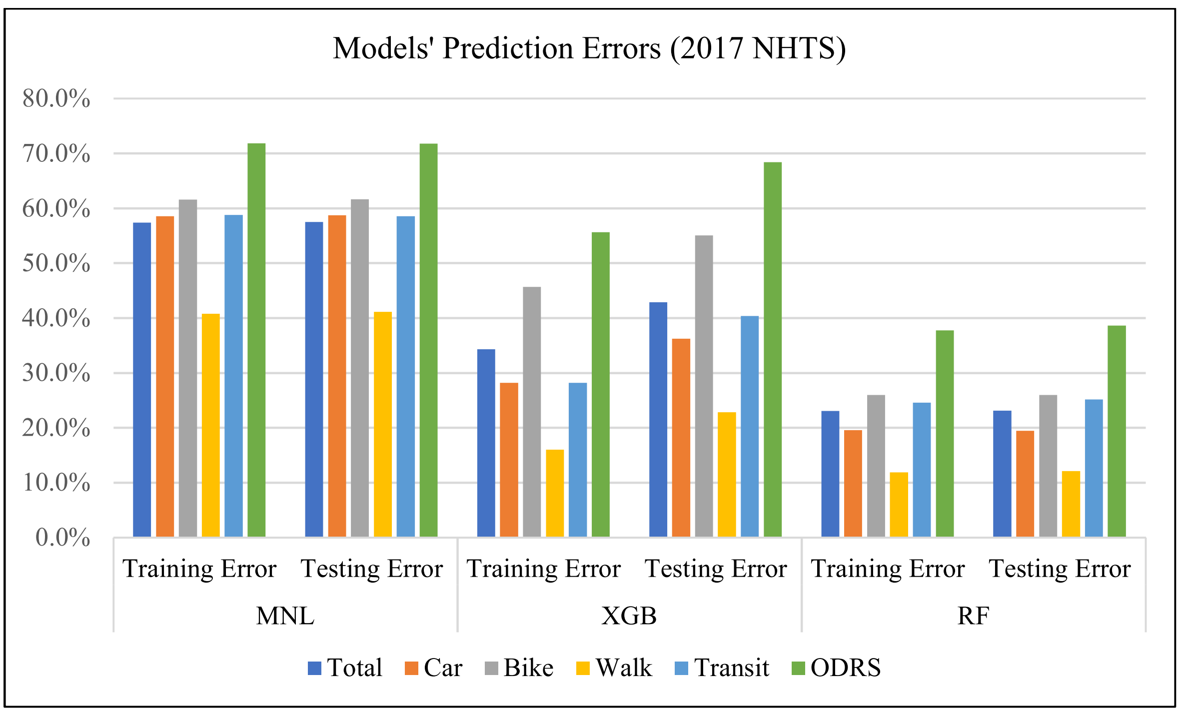


Figure 5.4. Models' Prediction Errors for the 2017 NHTS Data

The average training and testing errors of the three models for the New York region are presented in Table 5.9 and Figure 5.5 presents a quick comparison of the models' prediction errors. The main findings are fairly similar to those of the national model. First, though the three models all show good overall prediction accuracy, the two machine learning models' prediction errors are significantly lower than that of the MNL model. The XGB model has the highest overall prediction accuracy with a total training error of 4.1% and a total testing error of 10.0%. The RF model has a total training error of 9.6% and a total testing error of 8.9%. The MNL model has a total training error of 18.1% and a total testing error of 18.7%. Second, all the three models perform worst in predicting the choice of biking and ODRS as the two modes have the smallest shares. The two machine learning models perform slightly better than the MNL model in predicting the choice of ODRS, but perform significantly better in predicting the choice of biking. The XGB model has an ODRS-specific training error of 9.6% and an ODRS-specific testing error of 27.9%, while the ODRS-specific training and testing errors of the RF model are 27.3% and 26.7% respectively. The MNL model achieves an ODRS-specific training error of 25.3% and testing error of 28.6%. Third, the mode-specific prediction errors are not necessarily consistent with the magnitudes of the mode's share. All the three models perform worst in predicting the choice of ODRS and biking, but the models perform best in predicting the choice of car, even though it is not the mode with largest share in the dataset. The two machine learning models can achieve very low testing errors in predicting the choice of car, which are 2.6% and 3.9% respectively for the XGB model and the RF model. Though trips made by walking account for more than 50% in the dataset, and walking has the largest share, the three models' prediction errors of walking are higher than that of transit trips and trips by car. This is very interesting, as the models perform best in predicting the choice of walking in the national dataset, while perform much worse in predicting the choice of walking in the New York data. This could probably be associated with larger variation in characteristics of walking trips in the New York region. Fourth, the RF model and the MNL model are more consistent in training and testing errors indicating better generalizability. One commonly known drawback of machine learning model is the difficulty of avoiding the overfitting problem. The RF model is very good at avoiding the overfitting problem by design, but overfitting can be an issue for training the XGB model. In this analysis, the dataset is very unbalanced. The XGB model may not suffer from overfitting issue for all choices combined, but may have an overfitting issue when predicting the modes with small shares. For example, the hyper-parameters of the XGB models are tuned by minimizing the multi-class predicting error and the overfitting issue is controlled at the whole dataset level. However, since biking only accounts for 5% of the data, the tuned hyperparameters will likely result in overfitting for predicting biking choices. Therefore, machine learning techniques that can handle choice-specific overfitting control will be very helpful and needs to be further researched. Finally, the variances in the errors of the three models are very small, indicating that all three models are very robust to data changes.

Table 5.9. Models' Performance for the New York Region

New York		MNL		XGB		RF	
		Training Error	Testing Error	Training Error	Testing Error	Training Error	Testing Error
Total	Mean	18.1%	18.7%	4.1%	10.0%	9.6%	8.9%
	Variance	0.000006	0.000032	0.000506	0.000273	0.000017	0.000226
Car	Mean	8.8%	9.8%	1.1%	2.6%	3.8%	3.9%
	Variance	0.000007	0.000013	0.000050	0.000034	0.000002	0.000016
Bike	Mean	49.4%	53.1%	13.5%	30.3%	41.2%	40.8%
	Variance	0.000471	0.001101	0.009794	0.007203	0.000178	0.001123
Walk	Mean	29.7%	28.1%	7.1%	17.5%	10.2%	10.4%
	Variance	0.000022	0.000031	0.001076	0.000109	0.000011	0.000141
Transit	Mean	16.1%	15.4%	1.1%	5.8%	6.0%	5.9%
	Variance	0.000035	0.000435	0.000050	0.000140	0.000008	0.000134
ODRS	Mean	25.3%	28.6%	9.6%	27.9%	27.3%	26.7%
	Variance	0.000005	0.001159	0.003792	0.001732	0.000055	0.000645

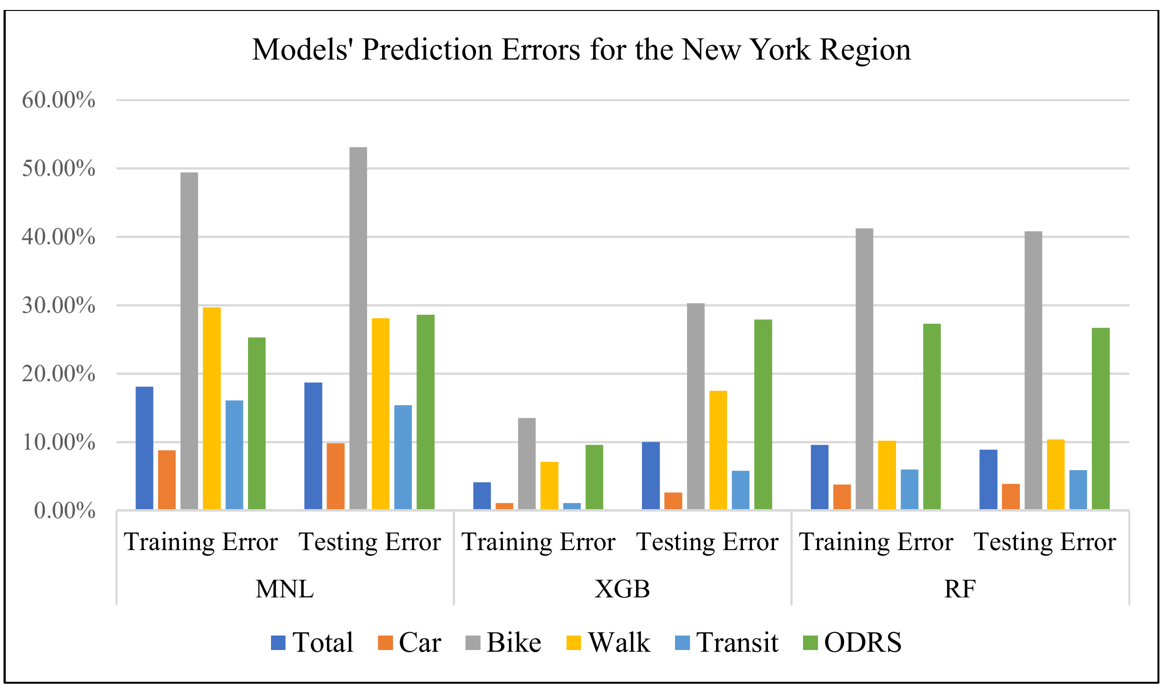


Figure 5.5. Models' Prediction Errors for the New York Region

The models' performance when applied to the Puget Sound region is tabulated in Table 5.10 and Figure 5.6 visualizes a quick comparison between the models' prediction errors. To a large extent, the models' performance is consistent with their performance for the New York data: (1) the two machine learning models are able to achieve lower total training and testing errors compared to the MNL model; (2) the three models perform best in predicting the choice of trips by car while perform poorly in predicting the choice of ODRS; (3) the RF model and the MNL model are more consistent in training and testing errors compared to the XGB model; and (4) all three models are very robust to data changes as the errors' variances are very small.

A main difference in the Puget Sound region's data is that it only contains 90 trips made by ODRS, while the New York data contains more than a thousand ODRS trips. Such a small number of observation of ODRS trips in the Puget Sound data results in extremely high prediction errors in predicting the choice of ODRS in all three models. Such a severe data limitation is hard to be overcome by any modeling techniques.

Table 5.10. Average Training and Testing Errors for Puget Sound Region

Puget Sound		MNL		XGB		RF	
		Training Error	Testing Error	Training Error	Testing Error	Training Error	Testing Error
Total	Mean	22.5%	22.8%	3.9%	17.5%	13.7%	14.2%
	Variance	0.000044	0.000134	0.000984	0.000382	0.003397	0.003107
Car	Mean	13.1%	13.1%	1.8%	8.4%	4.8%	4.8%
	Variance	0.000027	0.000132	0.000171	0.000115	0.000647	0.000625
Bike	Mean	36.2%	39.6%	8.1%	32.0%	50.6%	50.1%
	Variance	0.001002	0.004970	0.006038	0.010064	0.050019	0.051838
Walk	Mean	45.6%	46.5%	7.2%	33.7%	22.6%	22.7%
	Variance	0.000367	0.001227	0.003922	0.001088	0.008856	0.009486
Transit	Mean	30.0%	29.5%	1.8%	29.1%	32.7%	35.1%
	Variance	0.000547	0.002548	0.000171	0.002055	0.022918	0.028341
ODRS	Mean	80.1%	82.8%	33.0%	86.3%	91.8%	93.8%
	Variance	0.001520	0.007715	0.048570	0.005533	0.040626	0.027684

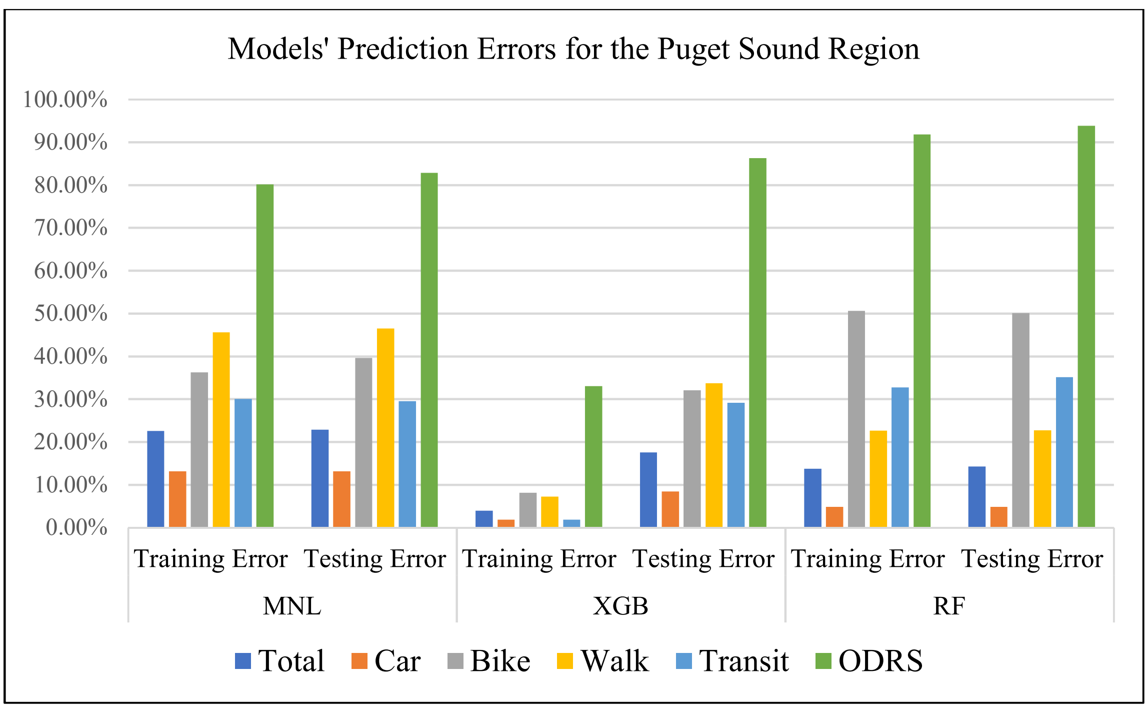


Figure 5.6. Models’ Prediction Errors for the Puget Sound Region

The models' performance when applied to the Delaware Valley region is tabulated in Table 5.11 and Figure 5.7 visualizes a quick comparison between the models' prediction errors. The three models have much higher prediction errors when applied to the Delaware Valley region's data. The RF model has the lowest total testing error of 27.9%, while the XGB model has a total testing error of 31.6% and the MNL models' total testing error is as high as 43.9%. Nevertheless, the main findings from the previous two regions still hold: (1) the two machine learning models are able to achieve lower total training and testing errors compared to the MNL model; (2) the three models perform best in predicting the choice of trips by car while perform worst in predicting the choice of ODRS; (3) the RF model and the MNL model are more consistent in training and testing errors compared to the XGB model; and (4) all three models are very robust to data changes as the errors' variances are very small. There are 190 observations of valid trips made by ODRS in the Delaware Valley region. Though the number is somewhat larger than that in the Puget Sound region's data, such a small number of observations makes it very hard to model the choices with many different independent variables.

Table 5.11. Average Training and Testing Errors for Delaware Valley Region

Delaware Valley		MNL		XGB		RF	
		Training Error	Testing Error	Training Error	Testing Error	Training Error	Testing Error
Total	Mean	43.3%	43.9%	11.0%	31.6%	28.2%	27.9%
	Variance	0.000055	0.000183	0.004749	0.000385	0.014563	0.015427
Car	Mean	23.2%	23.7%	3.1%	11.2%	11.5%	11.2%
	Variance	0.000069	0.000399	0.000437	0.000274	0.006907	0.007021
Bike	Mean	76.9%	77.7%	27.4%	79.2%	62.7%	63.1%
	Variance	0.000508	0.001286	0.045365	0.003919	0.074426	0.073116
Walk	Mean	60.2%	60.8%	17.4%	35.7%	30.2%	30.7%
	Variance	0.000387	0.001092	0.005498	0.000809	0.004285	0.004398
Transit	Mean	47.4%	47.9%	3.1%	39.4%	35.9%	35.8%
	Variance	0.000335	0.001152	0.000437	0.001119	0.038838	0.037994
ODRS	Mean	86.6%	86.2%	20.0%	87.2%	83.8%	84.6%
	Variance	0.000797	0.002343	0.033494	0.002456	0.046192	0.045305

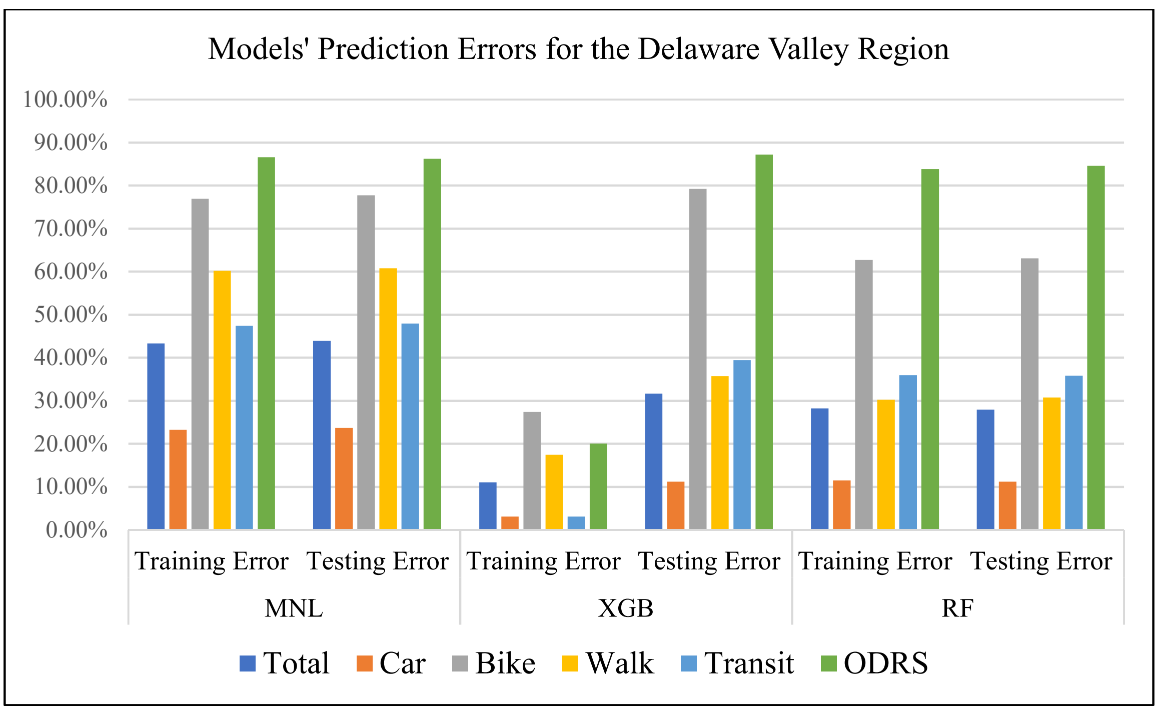


Figure 5.7. Models' Prediction Errors for the Delaware Valley Region

5.4 Models' Explanatory Results

5.4.1 Explanatory Results of the Models using the 2017 National Data

The MNL model developed using the 2017 NHTS data is shown in Table 5.12 and the MNL model developed using cluster-robust errors is presented in Table 5.13. The results of the two models are extremely similar and most of the estimates are the same or only slightly different. The significance levels of the variables are also almost the same in the two models. This is probably because that there is a large number of travelers in the data, so even if one traveler is associated with multiple trips, the issue of violation of independent observations is not a severe one.

The national model has a Pseudo R-squared of 0.32, which is decent for a 5-alternative model developed using a national dataset. According to the result of the national model, travel time, trip departure time, trip purpose, and number of travelers are the trip factors that are significantly associated with the choice of ODRS. Travel time is one of the most important determinants of travel mode choice. The coefficient of travel time for ODRS is -0.02, which is like the coefficient of travel time by car. ODRS trips are less likely to be in peak hours and are more likely to happen in late night. ODRS is more likely to be chosen for work, medical, and social trips and is less likely to be chosen for shopping trips. The choice of ODRS is positively associated with the total number of travelers in a trip.

Regarding socio-demographic characteristics, it is found that travelers from smaller households, households with no children, low-income households, and households with less vehicles, are more likely to choose ODRS. Travelers with higher education attainment are also positively associated with the choice of ODRS and elderly people are found to be negatively associated with the choice of ODRS. Travelers with a disability are found to be more likely to choose ODRS and transit. The choice of ODRS is also positively associated with high population density, which is probably a result of the fact that ODRS, especially ride-sourcing, is more accessible in those areas. People who use smartphones everyday are more likely to choose ODRS. Currently all ride-sourcing services are only available on smartphones, so the access to ODRS depends on using smartphones.

Table 5.12. MNL Model (2017 NHTS Data)

	Biking		Walking		Transit		ODRS		Car	
Variable	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**
Travel time	-0.04	***	-0.07	***	-0.003	***	-0.02	***	-0.02	***
(Intercept)	1.19	***	1.91	***	-0.57	***	-0.93	***		
Peak hour	0.002		-0.04		0.20	***	-0.31	***		
Late night	-0.60	***	-0.15		-0.18		1.50	***		
Loop trip	10.87	***	13.94	***	-16.70		-16.05			
Use smartphone everyday	-0.06		0.17	*	-0.06		0.31	***		
Trip purpose: work	-0.06		-0.09		0.52	***	0.20	*		
Trip purpose: medical	-0.82	**	-1.02	**	0.36	*	0.76	***		
Trip purpose: shopping	-0.95	***	-0.99	***	-0.81	***	-1.18	***		
Trip purpose: social	0.87	***	0.89	***	0.05		0.51	***		
Trip purpose: meals	-0.87	***	-0.35	**	-0.85	***	-0.02			
Number of travelers	-0.23	***	0.01		0.10	***	0.08	***		
Household size	-0.04		-0.04		-0.10	***	-0.10	**		
Low-income	0.88	***	0.50	***	1.23	***	0.75	***		
No children in the household	0.37	***	0.44	***	0.38	***	0.42	***		
Vehicle count per driver	-0.14	***	-0.17	***	-0.63	***	-0.44	***		
High population density	0.78	***	0.68	***	1.33	***	1.00	***		
Younger than 18	1.32	***	0.10		0.25	*	-0.10			
65+ years old	-1.06	***	-0.50	***	-0.55	***	-0.67	***		
Has bachelor degree or above	0.77	***	0.63	***	0.25	***	0.51	***		
Female	-1.03	***	-0.18	**	-0.19	***	-0.05			
Medical device used (disabilities)	-0.22	.	0.01		0.82	***	0.91	***		

Log-likelihood (equally-likely) = -29705.4; Log-likelihood (market-share) = -29540.7; Log-likelihood (full-model) = -20340.8

Pseudo R2 (equally-likely based) = 0.315

Table 5.13. MNL Model Using Cluster-Robust Standard Errors (2017 NHTS Data)

	Biking			Walking			Transit			ODRS			Car		
Variable	**Est.**	**t value**	**Sig.**	**Est.**	**t value**	**Sig.**	**Est.**	**t value**	**Sig.**	**Est.**	**t value**	**Sig.**	**Est.**	**t value**	**Sig.**
Travel time	-0.040	-16.6	***	-0.072	-25.0	***	-0.003	-9.3	***	-0.019	-11.0	***	-0.017	-10.4	***
(Intercept)	1.188	5.4	***	1.906	9.8	***	-0.566	-2.7	**	-0.929	-3.7	***			
Peak hour	0.002	0.0		-0.037	-0.6		0.196	3.6	***	-0.306	-4.9	***			
Late night	-0.601	-3.9	***	-0.153	-1.0		-0.178	-1.3		1.498	13.5	***			
Loop trip	10.866	5.0	***	13.935	6.3	***	-17.700	-26.1	***	-17.052	-27.9	***			
Use smartphone everyday	-0.062	-0.7		0.174	2.1	*	-0.059	-0.7		0.315	2.7	**			
Trip purpose: work	-0.057	-0.6		-0.090	-0.9		0.520	7.1	***	0.199	2.2	*			
Trip purpose: medical	-0.820	-3.3	***	-1.022	-3.4	***	0.364	2.1	*	0.755	4.5	***			
Trip purpose: shopping	-0.951	-12.1	***	-0.993	-12.1	***	-0.808	-10.4	***	-1.182	-10.4	***			
Trip purpose: social	0.872	10.3	***	0.890	9.9	***	0.054	0.6		0.514	6.0	***			
Trip purpose: meals	-0.871	-7.0	***	-0.350	-3.4	***	-0.847	-7.6	***	-0.016	-0.2				
Number of travelers	-0.231	-2.3	*	0.010	0.4		0.096	5.9	***	0.085	5.8	***			
Household size	-0.038	-1.0		-0.043	-1.2		-0.097	-3.0	**	-0.098	-2.1	*			
Low-income	0.881	9.1	***	0.500	5.2	***	1.232	11.2	***	0.751	6.1	***			
No children in the household	0.374	3.6	***	0.445	4.3	***	0.382	4.1	***	0.416	3.7	***			
Vehicle count per driver	-0.135	-3.6	***	-0.169	-3.9	***	-0.633	-4.1	***	-0.443	-4.0	***			
High population density	0.779	11.8	***	0.675	10.7	***	1.332	20.4	***	1.001	13.2	***			
Younger than 18	1.322	10.8	***	0.099	0.8		0.248	2.0	*	-0.096	-0.5				
65+ years old	-1.058	-11.2	***	-0.497	-6.0	***	-0.547	-6.7	***	-0.675	-6.5	***			
Has bachelor degree or above	0.769	9.9	***	0.630	8.9	***	0.251	3.9	***	0.506	6.0	***			
Female	-1.032	-15.8	***	-0.175	-2.8	**	-0.188	-3.3	**	-0.047	-0.6				
Medical device used (disabilities)	-0.219	-1.4		0.014	0.1		0.820	6.9	***	0.906	6.3	***			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Both the RF and XGB models are tree-based models that allow the measurement of the “importance” a variable has in forming the final rule-based classifications. The importance measures used by the RF models and the XGB models are slightly different, as the RF model uses “mean decrease in accuracy”, while the XGB model uses a composite index called “importance”.

Figure 5.8 and Figure 5.9 plot the top 15 important variables for the RF and XGB models for the national data. Travel time, trip distance, travel speed, and trip cost are found to be the most important variables influencing people’s mode choices. The four variables together explain more than 40% in the RF model and more than 70% of trip mode prediction in the XGB model. Vehicle ownership, loop trips, number of travelers, population density, and so on are also among the most important variables. The independent variables that are important in the RF and XGB models are very consistent and most of them are also included in the developed MNL model.

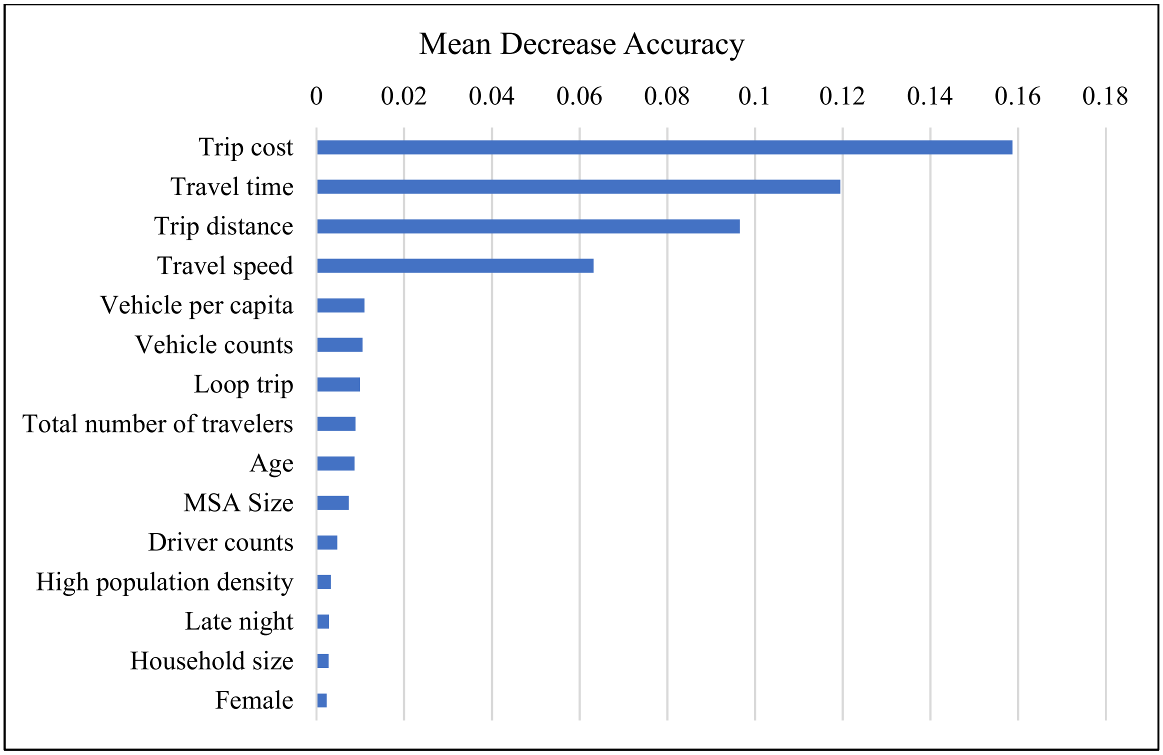


Figure 5.8. Top 15 Important Variables of the RF Model (2017 NHTS)

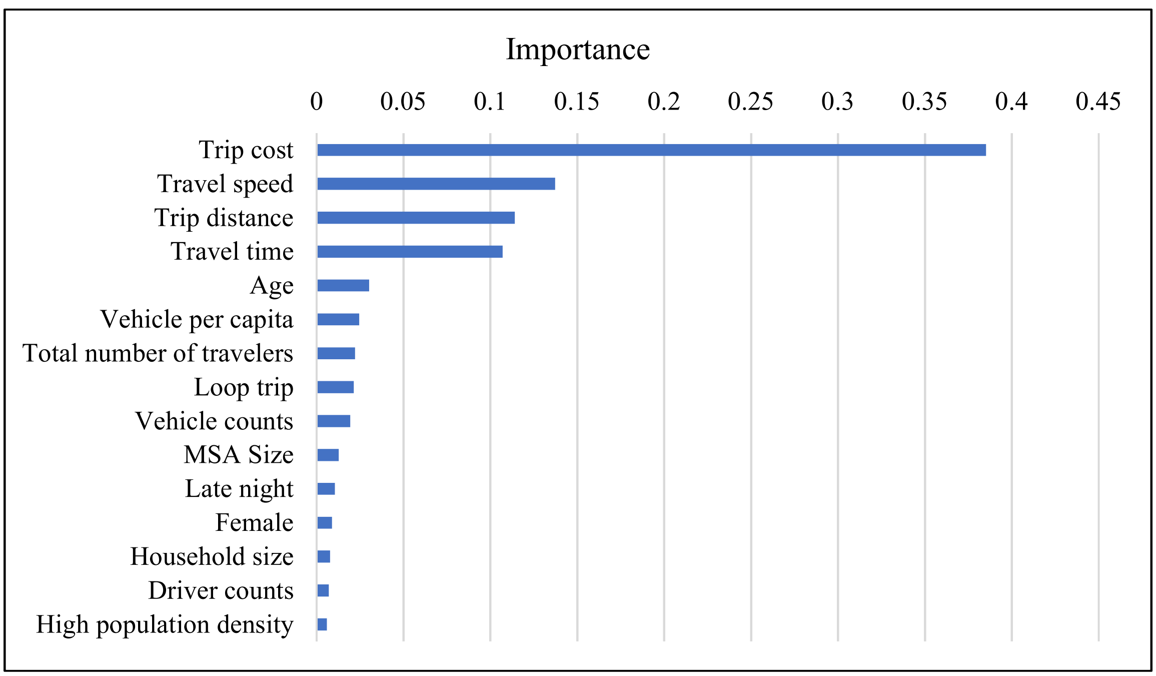


Figure 5.9. Top 15 Important Variables of the XGB Model (2017 NHTS)

5.4.2 *Explanatory Results of the Models for the New York Region*

The estimation result of the MNL model for the New York region is presented in Table 5.14 and the MNL model using cluster-robust standard errors is presented in Table 5.15. Similar to the previous MNL models using the 2017 NHTS data, the results of the two models using the New York data are also very similar and most of the estimates are the same or only slightly different. The significance levels of the variables are also almost the same in the two models.

Overall, the model has an excellent goodness of fit, as the adjusted Rho-squared (McFadden R-squared) of this model is about 0.767, indicating that about 76.7% of the information contained in the data is explained by this model. Most of the variables included in the model were set as individual-specific variables, indicating that the values of the variables only vary by the traveler or the trip, but do not vary by travel alternatives. The variable of travel time is specified as an alternative-specific variable, as its value varies by each alternative of a certain trip. The signs of all the variables included in this final model are consistent with theory and all the three types of variables (the trip variables, the personal/household variables, and the neighborhood variables) turned out to be significantly associated with people's travel mode choices.

The travel mode of car is set as the default mode in the model, and the other four modes all have negative intercepts. The ODRS has an intercept of -2.91 whose magnitude is only smaller than that of the transit-specific intercept which is -6.85, indicating that when all other factors are the same, a traveler is less likely to choose transit and ODRS compare to biking and walking.

The variable of ODRS-specific travel time has a coefficient of -0.09, the magnitude of which is larger than the choice of car (-0.08) and transit (-0.08), like the biking- specific coefficient that is -0.09, and smaller than the walking-specific coefficients that are -0.15. This indicates for longer trips that people are more likely to choose driving and biking than ODRS, but tend to prefer ODRS to walking and transit to some degree, keeping all other conditions the same. Trip purpose is also significantly associated with the choice of ODRS. Among the five types of trip purposes, the purpose of changing travel mode has the largest positive coefficient for the choice of ODRS. The trip purpose of personal/household maintenance work, which are trips made for personal services, appointments, and shopping needed by the individual or household, is also statistically significant and positively associated with the choice of ODRS. The variables of morning-peak and evening-peak have negative signs and the variable of late night has positive sign for the choice of ODRS, suggesting that people are more likely to choose ODRS for late night trips rather than peak hour trips. The variable of activity duration has a positive sign for the choice of ODRS and walking, while it has a negative sign for the choice of biking and transit.

Many of the personal and household characteristics included in the model are significantly associated with the choice of ODRS. Both low income and high-income travelers are positively associated with ODRS trips. Disabled people and female travelers are also positively associated with the choice of ODRS. Travelers from smaller households are more likely to choose ODRS than larger households. Students and households with higher number of vehicles per capita are negatively associated with ODRS.

Multiple neighborhood variables are found to be significantly associated with the choice of ODRS and all the neighborhood variables included in the model are significantly and positively associated with the choice of ODRS, except the variable of employment diversity at the trip destination. Generally, these neighborhood variables are density and diversity measures and their positive signs reflect that ODRS trips are more likely to happen in more densely developed areas with mixed land use development patterns, denser roads, and higher transit service levels.

Table 5.14. MNL Model (New York Region)

Variable	Biking		Walking		Transit		ODRS		Car	
	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.
Travel time	-0.090	***	-0.151	***	-0.084	***	-0.091	***	-0.075	***
(Intercept)	0.318		-0.452	*	-7.038	***	-3.086	***		
Purpose: work	0.078		0.284	*	2.173	**	0.585	**		
Purpose: recreation	-0.058		0.319	**	0.070		0.132			
Purpose: maintenance	0.371	*	0.331	**	0.025		0.514	**		
Purpose: change mode	0.742	**	2.908	***	6.664	***	2.471	***		
Morning peak	-0.189		0.0053		0.017		-0.642	***		
Evening peak	0.353	**	0.275	**	0.218		-0.104			
Late night	-0.463		-0.053		-0.398		0.400	.		
Activity duration	0.0003		0.0011	***	-0.0057	***	0.0013	***		
Number of travelers	-0.577	***	-0.429	***	-0.150	*	-0.081	.		
Low-income	-0.242		0.514	***	0.502	**	0.715	***		
High-income	0.302	*	-0.141		-0.302	.	0.549	***		
65+ years old	-1.174	***	-0.335	*	-0.050		0.193			
Disability	-0.262		-0.228		0.289		0.883	***		
Female	-0.722	***	-0.058		-0.160		0.301	**		
Household size	-0.328	***	-0.191	***	-0.227	***	-0.441	***		
Student	0.625	***	0.926	***	1.022	***	-0.479	.		
Vehicle count per driver	-1.691	***	-1.814	***	-2.058	***	-2.341	***		
O: population density	9.20E-06	***	1.18E-05	***	1.25E-05	***	1.44E-05	***		
O: employment density	4.18E-07		2.85E-06	***	3.06E-06	***	3.35E-06	***		
O: subway density	3.49E-02	***	3.48E-02	***	6.16E-02	***	2.66E-02	**		
O: road density	1.04E-05	*	1.83E-05	***	1.81E-05	***	1.50E-05	***		
D: population density	5.96E-06	*	9.87E-06	***	1.48E-05	***	1.03E-05	***		
D: subway density	1.91E-02	*	1.85E-02	*	1.88E-02	*	1.74E-02	*		
D: road density	1.06E-05	*	2.01E-05	***	1.44E-05	**	2.02E-05	***		

Note: “O” stands for trip origin and “D” stands for trip destination.

Log-likelihood (equally-likely) = -25751.0; Log-likelihood (market-share) = -20620.2; Log-likelihood (full-model) = -6000.4

Pseudo R2 (equally-likely based) = 0.767

Table 5.15. MNL Model Using Cluster-Robust Standard Errors (New York Region)

Variable	Biking			Walking			Transit			ODRS			Car		
	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.
Travel time	-0.090	-14.0	***	-0.151	-19.9	***	-0.084	-10.0	***	-0.091	-9.1	***	-0.075	-7.5	***
(Intercept)	0.318	0.9		-0.452	-1.7	.	-7.038	-8.1	***	-3.086	-9.6	***			
Purpose: work	0.078	0.4		0.284	2.1	*	2.173	2.8	**	0.585	3.6	***			
Purpose: recreation	-0.058	-0.4		0.319	2.7	**	0.070	0.1		0.132	0.8				
Purpose: maintenance	0.371	2.3	*	0.331	2.7	**	0.025	0.0		0.514	3.4	***			
Purpose: change mode	0.742	3.1	**	2.908	18.1	***	6.664	9.5	***	2.471	12.5	***			
Morning peak	-0.189	-1.4		0.0053	0.1		0.017	0.1		-0.642	-5.0	***			
Evening peak	0.353	2.8	**	0.275	3.0	**	0.218	1.4		-0.104	-0.8				
Late night	-0.463	-1.9	.	-0.053	-0.3		-0.398	-1.3		0.400	1.9	.			
Activity duration	0.0003	1.0		0.0011	6.0	***	-0.0057	-2.7	**	0.0013	5.2	***			
Number of travelers	-0.577	-5.5	***	-0.429	-4.8	***	-0.150	-0.8		-0.081	-1.4				
Low-income	-0.242	-1.5		0.514	4.7	***	0.502	2.8	**	0.715	5.3	***			
High-income	0.302	2.6	**	-0.141	-1.6		-0.302	-2.0	*	0.549	4.9	***			
65+ years old	-1.174	-5.8	***	-0.335	-2.5	*	-0.050	-0.2		0.193	1.4				
Disability	-0.262	-1.1		-0.228	-1.5		0.289	1.1		0.883	5.8	***			
Female	-0.722	-7.1	***	-0.058	-0.8		-0.160	-1.3		0.301	3.1	**			
Household size	-0.328	-7.3	***	-0.191	-5.6	***	-0.227	-3.9	***	-0.441	-9.7	***			
Student	0.625	3.8	***	0.926	7.5	***	1.022	4.7	***	-0.479	-2.2	*			
Vehicle count per driver	-1.691	-8.6	***	-1.814	-13.1	***	-2.058	-8.5	***	-2.341	-12.2	***			
O: population density	9.20E-06	4.7	***	1.18E-05	7.1	***	1.25E-05	5.7	***	1.44E-05	8.9	***			
O: employment density	4.18E-07	0.8		2.85E-06	7.0	***	3.06E-06	6.8	***	3.35E-06	7.7	***			
O: subway density	3.49E-02	4.9	***	3.48E-02	6.1	***	6.16E-02	8.8	***	2.66E-02	3.9	***			
O: road density	1.04E-05	2.5	*	1.83E-05	5.1	***	1.81E-05	3.9	***	1.50E-05	4.8	***			
D: population density	5.96E-06	3.1	**	9.87E-06	5.9	***	1.48E-05	6.8	***	1.03E-05	6.1	***			
D: subway density	1.91E-02	2.4	*	1.85E-02	2.8	**	1.88E-02	2.5	*	1.74E-02	2.4	*			
D: road density	1.06E-05	2.5	*	2.01E-05	5.9	***	1.44E-05	3.2	**	2.02E-05	6.4	***			

Note: “O” stands for trip origin and “D” stands for trip destination.

Both the RF and XGB models are tree-based models that allowed the measurement of the "importance" a variable in forming the final rule-based classifications. The importance measures used by the RF models and the XGB models are slightly different, as the RF model uses "mean decrease in accuracy", while the XGB model uses a composite index called "importance".

Figure 5.10 and Figure 5.11 plot the 15 most important variables for the RF and XGB models for the New York region respectively. Travel time, trip distance, and trip cost are found to be the most important variables influencing people's mode choices. The three variables together explain more than 50% of trip mode predicting in both the RF and the XGB model. Travel time and vehicle ownership are also among the top important variables. Many neighborhood variables are found to be among the top important variables.

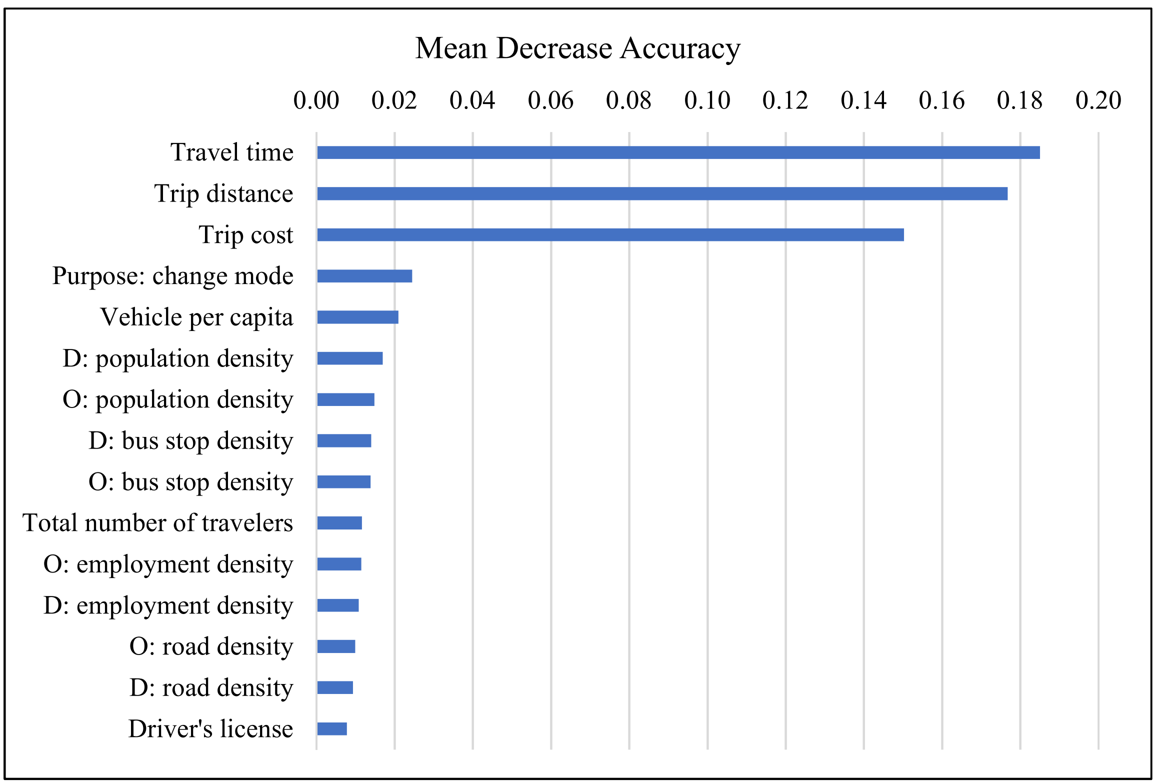


Figure 5.10. Top 15 Important Variables of the RF Model (New York Region)

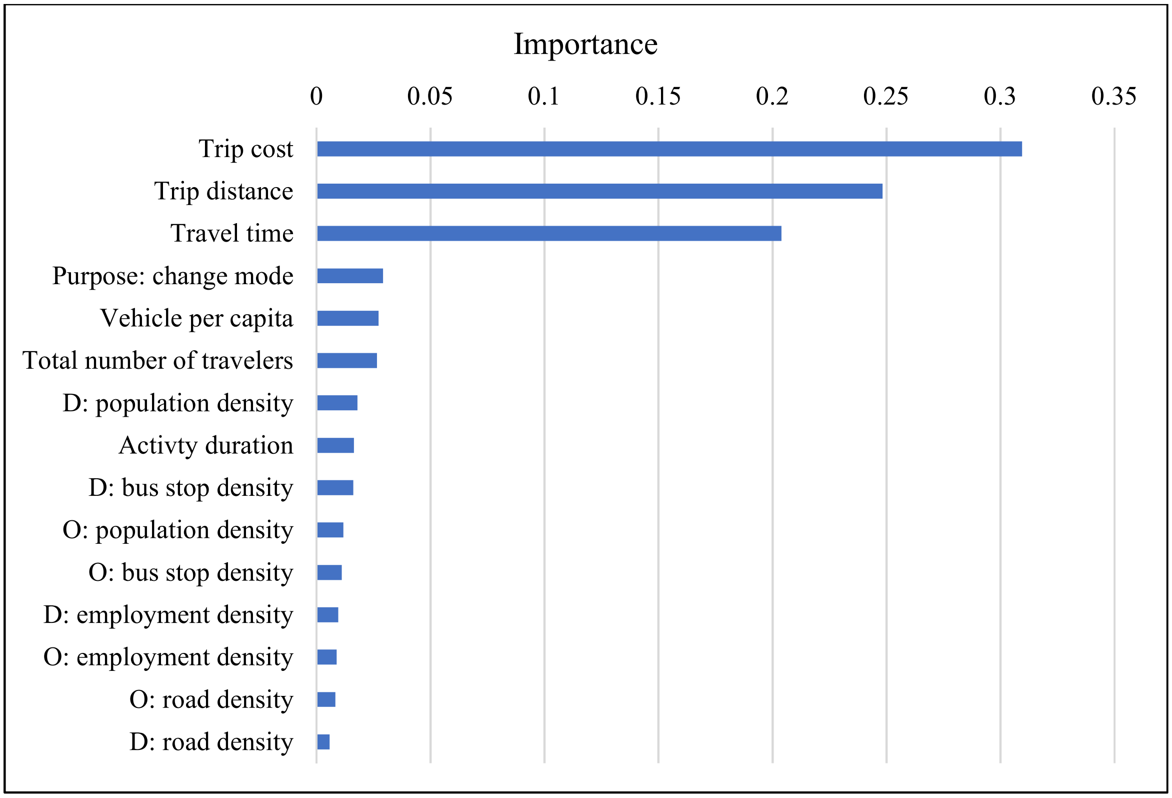


Figure 5.11. Top 15 Important Variables the XGB Model (New York Region)

5.4.3 Explanatory Results of the Models for the Puget Sound Region

The estimation result of the MNL model developed for the Puget Sound region is presented in Table 5.16 and the MNL model developed using cluster-robust errors is presented in Table 5.17. The results of the two models are extremely similar and most of the estimates are the same or only slightly different. The significance levels of the variables are also almost the same in the two models. This model has an overall goodness of fit, measured by the adjusted Rho-squared (McFadden R-squared) as 0.72, which is very high and is like that of the New York model. The three types of variables, including the trip characteristics, the personal/household variables, and the neighborhood variables are all found to be significantly associated with people's mode choices.

Probably due to the extremely small number of trips made by ODRS in the Puget Sound region, much fewer independent variables are statistically significant associated with the choice of ODRS. The choice of ODRS is also positively associated with trips made for changing travel modes, indicating that ODRS serves more multimodal travel. ODRS trips are also strongly associated with the dummy variable of late night, indicating that for late night trips, people are more likely to choose ODRS, which is consistent with the finding of the New York's model.

Household size and vehicle ownership are the only two personal/household variables found to be significantly associated with the choice of ODRS. Travelers from smaller households with less vehicle ownership are positively associated with the choice of ODRS, which is consistent with the findings from of the New York model. Population density, employment density, and job-housing balance are found to be positively associated with the choice of ODRS, which is probably associated with the supply level of ODRS.

Table 5.16. MNL Model (Puget Sound Region)

	Biking		Walking		Transit		ODRS		Car	
	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.
Trip cost (generic variable)	Est.: - 0.131 Sig.: ***									
Travel time	-0.083	***	-0.124	***	-0.059	***	-0.192	***	-0.346	***
(Intercept)	0.754		1.735	***	0.460		-1.850	*		
Purpose: maintenance	-1.101	***	-0.815	***	-0.758	**	-0.081			
Purpose: recreation	0.107		0.575	***	-0.294		0.150			
Purpose: change mode	5.696	*	5.987	**	-7.546		3.640	***		
Number of travelers	-0.345	*	-0.051		-0.029		-0.258			
Late night	-0.561		-0.671	*	-0.434		1.799	***		
Vehicle count per cap	-1.454	***	-1.488	***	-1.948	***	-2.905	***		
Household size	-0.090		-0.234	***	-0.311	**	-0.412	*		
Low-income	0.313		-0.045		0.511	.	0.115			
65+ years old	-2.202	***	-0.761	***	-0.642	*	0.404			
Driver's license	-0.316		-0.066		-0.558	*	-0.408			
Female	-0.821	***	-0.300	*	-0.351	.	0.011			
O: bus stop density	6.22E-04		1.41E-03	**	2.20E-03	***	1.97E-03	**		
O: job housing balance	1.544	***	1.317	***	1.137	**	2.562	***		
D: bus stop density	-3.88E-04		1.14E-03	*	1.86E-03	***	2.23E-03	**		

Note: "O" stands for trip origin and "D" stands for trip destination.

Log-likelihood (equally-likely) = -7242.5; Log-likelihood (market-share) = -4374.0; Log-likelihood (full-model) = -2027.6

Pseudo R2 (equally-likely based) = 0.720

Table 5.17. MNL Model Using Cluster-Robust Standard Errors (Puget Sound)

Variable	Biking			Walking			Transit			ODRS			Car		
	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.
Trip cost (generic variable)	Est.: - 0.131 Sig.: **														
Travel time	-0.083	-15.4	***	-0.124	-20.5	***	-0.059	-17.4	***	-0.192	-4.6	***	-0.346	-10.1	***
(Intercept)	0.754	1.1		1.735	3.9	***	0.460	0.8		-1.850	-1.4				
Purpose: maintenance	-1.101	-4.1	***	-0.815	-4.4	***	-0.758	-3.1	**	-0.081	-0.2				
Purpose: recreation	0.107	0.5		0.575	3.8	***	-0.294	-1.5		0.150	0.4				
Purpose: change mode	5.696	5.8	***	5.987	6.2	***	-7.546	-6.1	***	3.640	4.7	***			
Number of travelers	-0.345	-2.4	*	-0.051	-0.6		-0.029	-0.2		-0.258	-1.0				
Late night	-0.561	-1.6		-0.671	-2.3	*	-0.434	-1.2		1.799	5.1	***			
Vehicle count per cap	-1.454	-5.6	***	-1.488	-8.4	***	-1.948	-9.4	***	-2.905	-6.6	***			
Household size	-0.090	-1.0		-0.234	-3.2	**	-0.311	-3.4	***	-0.412	-1.9	.			
Low-income	0.313	0.8		-0.045	-0.2		0.511	1.8	.	0.115	0.2				
65+ years old	-2.202	-4.8	***	-0.761	-3.6	***	-0.642	-2.4	*	0.404	1.0				
Driver's license	-0.316	-1.0		-0.066	-0.3		-0.558	-2.2	*	-0.408	-1.0				
Female	-0.821	-4.2	***	-0.300	-2.2	*	-0.351	-2.2	*	0.011	0.0				
O: bus stop density	6.22E-04	1.0		1.41E-03	2.9	**	2.20E-03	4.5	***	1.97E-03	3.1	**			
O: job housing balance	1.544	3.7	***	1.317	4.5	***	1.137	3.4	***	2.561	2.7	**			
D: bus stop density	-3.88E-04	-0.6		1.14E-03	2.2	*	1.86E-03	3.7	***	2.23E-03	3.4	***			

Note: “O” stands for trip origin and “D” stands for trip destination.

Figure 5.12 and Figure 5.13 plot the 15 most important variables for the RF and XGB models for the Puget Sound region respectively. Consistent with the New York model, travel time, trip cost, and trip distance are found to be the most important variables influencing people's mode choices in both the RF and the XGB models. The three variables explain more than 50% of the mode choices predicted by the models. Total number of travelers, vehicle ownership, whether the traveler has a driver's license, and activity duration are found to be among the most important factors, which are also consistent with the models for New York. Multiple neighborhood variables, including employment-population balance, population density, bus stop density, and employment density are found to be very important for predicting people's mode choices.

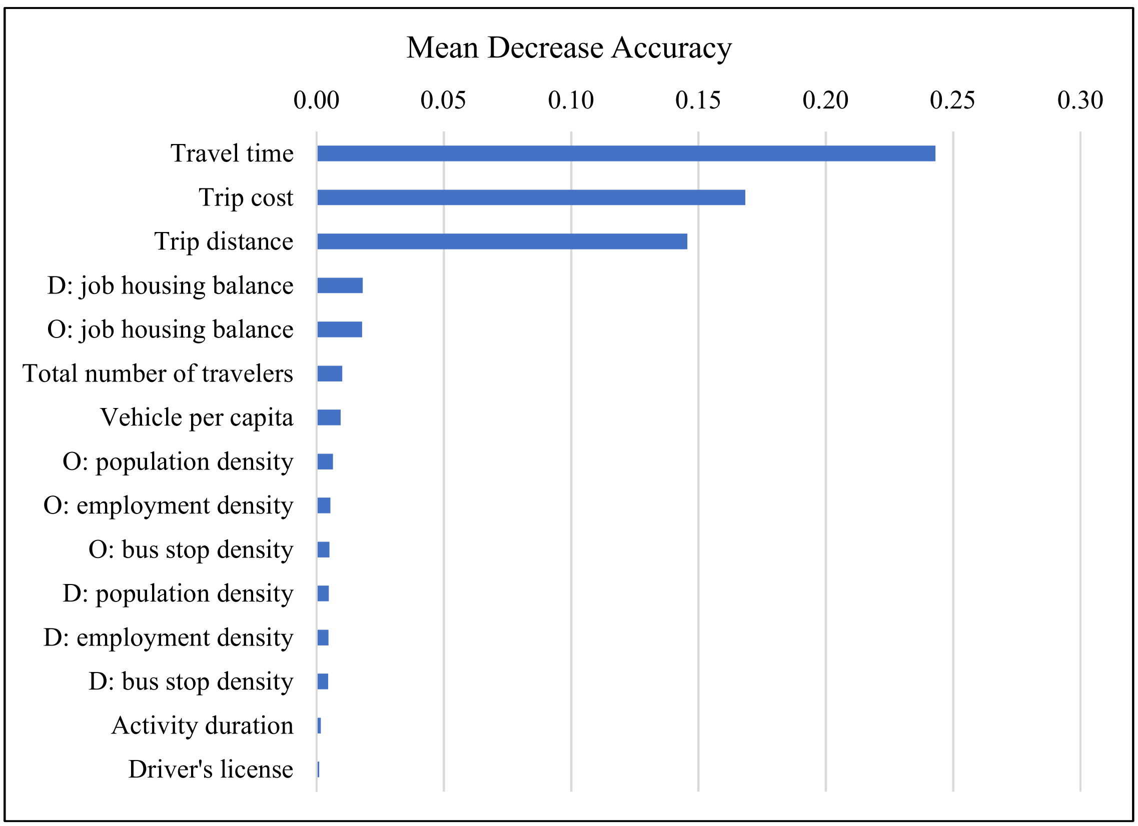


Figure 5.12. Top 15 Important Variables of the RF Model (Puget Sound Region)

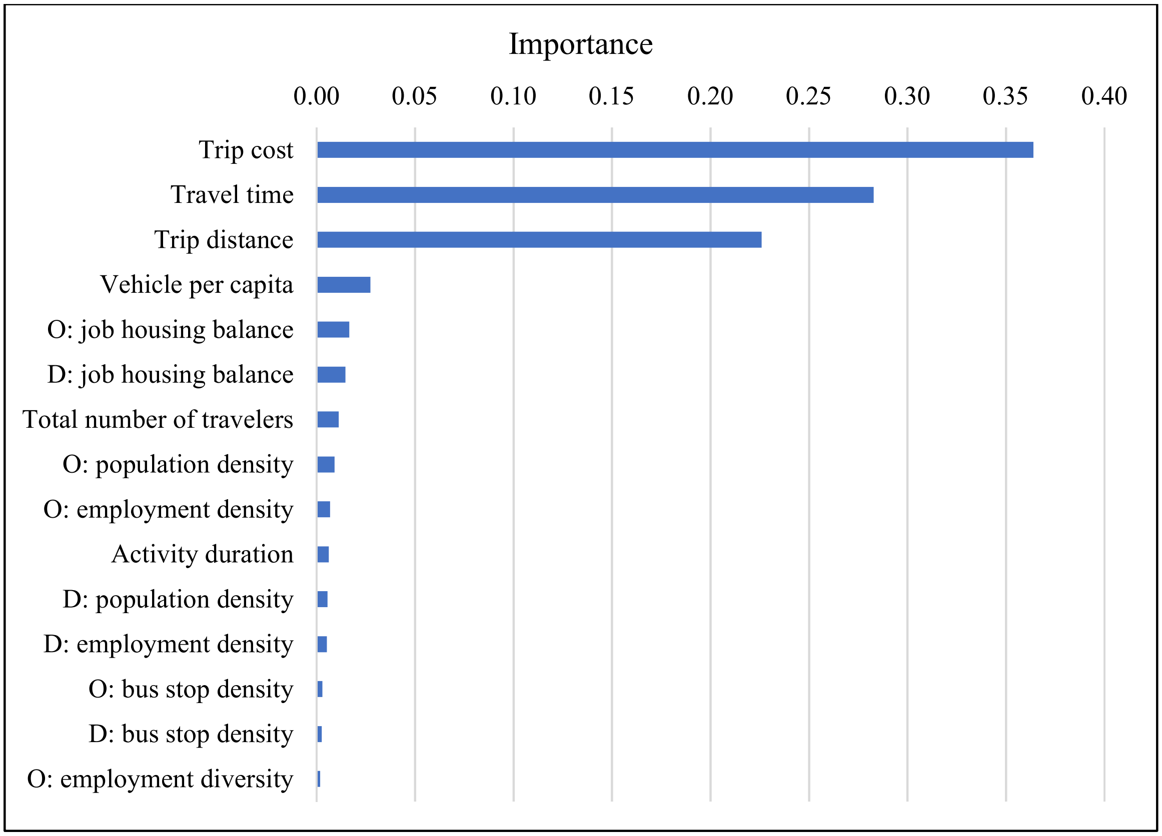


Figure 5.13. Top 15 Important Variables of the XGB Model (Puget Sound)

5.4.4 Explanatory Results of the Models for the Delaware Valley Region

The estimation result of the MNL model developed for the Delaware Valley region is presented in Table 5.18 and the MNL model developed using cluster-robust errors is presented in Table 5.19. Similar to previous MNL models, the results of the two models are very similar and most of the estimates are the same or only slightly different. The significance levels of the variables are also almost the same in the two models. This model has an overall goodness of fit, measured by the adjusted Rho-squared (McFadden R-squared) as 0.52, which is bigger than the Rho-squared value of the national model but a lot smaller than the Rho-squared values of the Puget Sound model or the New York model.

The three types of variables (the trip characteristics, the personal/household variables, and the neighborhood variables) are all found to be significantly associated with people's mode choices.

The number of trips made by ODRS in the Delaware Valley is also very small, but more independent variables are found to be significantly associated with the choice of ODRS in the Delaware Valley region compared to the Puget Sound region. The choice of ODRS is negatively associated with home-based other trips. ODRS trips are also strongly associated with the dummy variable of late night, indicating that for late night trips, people are more likely to choose ODRS, which is consistent with the finding of both the New York and the Puget Sound model.

Household size, vehicle ownership, and having a driver's license are found to be negatively and significantly associated with the choice of ODRS. Travelers from smaller households, travelers with less vehicle ownership, and travelers without driver's license, are more likely to choose ODRS, which is again consistent with the findings from of the New York and the Puget Sound models. Travelers younger than 18 years old are less likely to use ODRS, while travelers who are elderly than 65 years old are more like to use ODRS. Disability is also found to be significantly and positively associated with the choice of ODRS, which is consistent with the findings from the New York and the national models that ODRS disproportionately serves disabled people. If the traveler must pay for parking at the destination, the traveler is more likely to use ODRS. Population density is found to be positively associated with the choice ODRS, which is probably associated with the supply level of ODRS.

Table 5.18. MNL Model (Delaware Valley Region)

Variable	Biking		Walking		Transit		ODRS		Car	
	Est	Sig.	Est	Sig.	Est	Sig.	Est	Sig.	Est	Sig.
Travel time (generic variable)	Est.: - 0.026 Sig.: ***									
Trip cost	-2.470	***	-0.632	***	-0.362	***	-0.037	***	-0.167	***
(Intercept)	2.614	***	2.994	***	3.902	***	1.861	***		
Home-based other	-0.530	***	-0.152		-0.701	***	-0.781	***		
Peak hour	0.210	.	-0.028		0.340	**	-0.398	*		
Late night	-0.007		-0.265		-0.226		0.756	*		
Household size	-0.319	***	-0.292	***	-0.357	***	-0.420	***		
Vehicle per capita	-2.043	***	-1.473	***	-1.811	***	-2.055	***		
Low income	-0.186		0.348	*	0.592	***	-0.492	.		
High income	-0.305	*	-0.057		-0.342	**	0.157			
Female	-0.966	***	-0.271	**	-0.201	.	0.011			
Younger than 18	-0.338		-0.662	**	-1.150	***	-0.931	*		
65+ years old	-0.676	***	-0.421	**	0.069		0.761	***		
Have driver's license	-1.168	***	-1.751	***	-2.360	***	-2.497	***		
Disability	-0.568		-0.234		0.304		2.064	***		
Must pay for parking	0.945	***	0.714	***	0.748	***	0.524	*		
Employer has transit subsidy	1.016	***	0.509	*	0.530	**	0.084			
O: population density	7.49E-06	***	8.79E-06	***	7.70E-06	***	8.72E-06	***		
D: population density	1.07E-05	***	1.17E-05	***	1.17E-05	***	1.06E-05	***		

Note: “O” stands for trip origin and “D” stands for trip destination.

Log-likelihood (equally-likely) = -8047.2; Log-likelihood (market-share) = -6791.9; Log-likelihood (full-model) = -4505.9

Pseudo R2 (equally-likely based) = 0.440

Table 5.19. MNL Model Using Cluster-Robust Standard Errors (Delaware Valley)

Variable	Biking			Walking			Transit			ODRS			Car		
	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.	Est.	t value	Sig.
Travel time (generic variable)	Est.: - 0.026 t value: - 7.6 Sig.: ***														
Trip cost	-2.470	-7.9	***	-0.632	-4.345	***	-0.362	-4.572	***	-0.037	-4.601	***	-0.167	-9.4	***
(Intercept)	2.614	5.2	***	2.994	8.5	***	3.902	10.0	***	1.861	3.5	***			
Home-based other	-0.530	-2.9	**	-0.152	-1.244		-0.701	-5.074	***	-0.781	-2.909	**			
Peak hour	0.210	1.5		-0.028	-0.3		0.340	3.1	**	-0.398	-2.2	*			
Late night	-0.007	0.0		-0.265	-1.0		-0.226	-0.9		0.756	2.4	*			
Household size	-0.319	-3.8	***	-0.292	-5.8	***	-0.357	-6.3	***	-0.420	-3.6	***			
Vehicle per capita	-2.043	-6.7	***	-1.473	-8.4	***	-1.811	-9.6	***	-2.055	-6.4	***			
Low income	-0.186	-0.6		0.348	2.3	*	0.592	3.6	***	-0.492	-1.5				
High income	-0.305	-1.5		-0.057	-0.5		-0.342	-2.4	*	0.157	0.6				
Female	-0.966	-5.6	***	-0.271	-2.6	**	-0.201	-1.7	.	0.011	0.1				
Younger than 18	-0.338	-0.8		-0.662	-2.6	*	-1.150	-3.9	***	-0.931	-1.8	.			
65+ years old	-0.676	-2.4	*	-0.421	-2.7	**	0.069	0.4		0.761	2.7	**			
Have driver's license	-1.168	-3.0	**	-1.751	-7.4	***	-2.360	-9.5	***	-2.497	-7.1	***			
Disability	-0.568	-1.0		-0.234	-0.7		0.304	0.9		2.064	5.0	***			
Must pay for parking	0.945	3.8	***	0.714	4.2	***	0.748	4.2	***	0.524	1.8	.			
Employer has transit subsidy	1.016	3.5	***	0.509	2.1	*	0.530	2.2	*	0.084	0.2				
O: population density	7.49E-06	5.3	***	8.79E-06	6.3	***	7.70E-06	5.6	***	8.72E-06	5.9	***			
D: population density	1.07E-05	6.0	***	1.17E-05	6.6	***	1.17E-05	6.6	***	1.06E-05	4.9	***			

Note: “O” stands for trip origin and “D” stands for trip destination.

Figure 5.14 and Figure 5.15 plot the top 15 important variables for the RF and XGB models for the Delaware Valley region respectively. Similar to the findings from the previous models, travel time, trip cost, and trip distance are found to be the most important variables. Also, total number of travelers, vehicle ownership, whether the traveler has a driver’s license, and activity duration are found to be among the most important factors. Multiple neighborhood variables, including employment-population balance, population density, bus stop density, and employment density are found to be very important for influencing people’s mode choices.

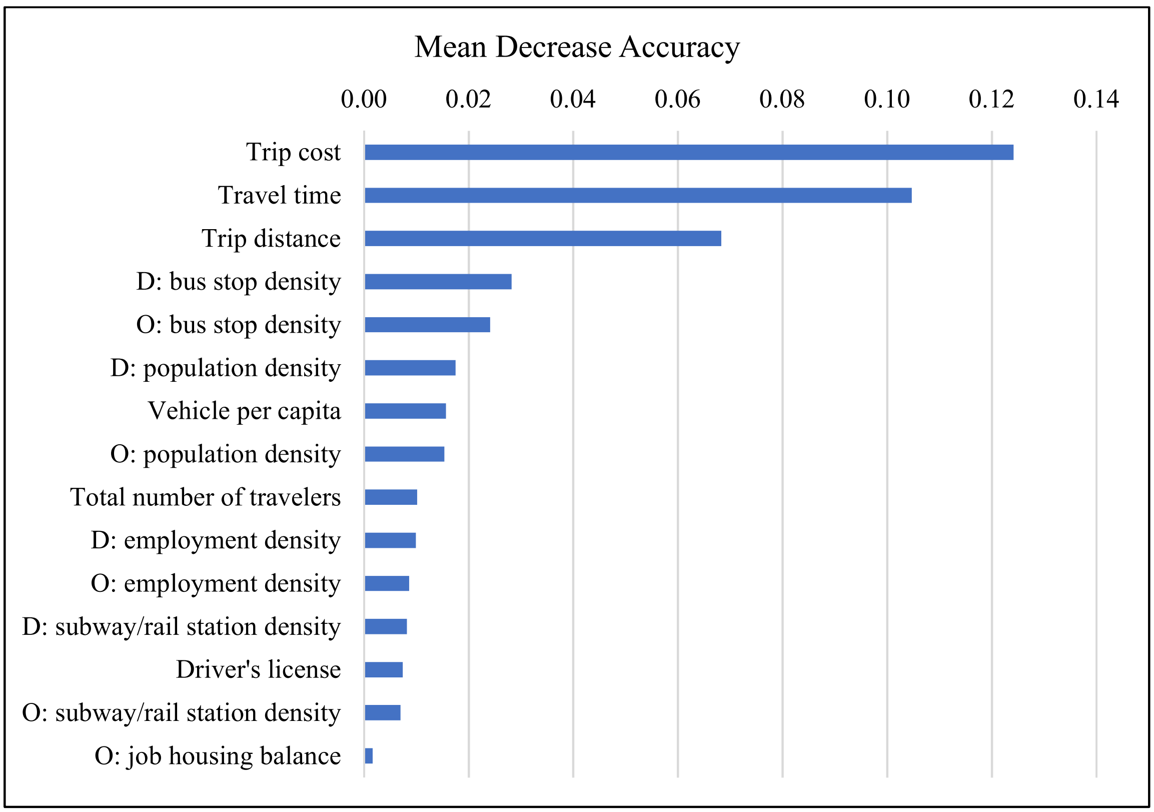


Figure 5.14. Top 15 Important Variables of the RF Model (Delaware Valley)

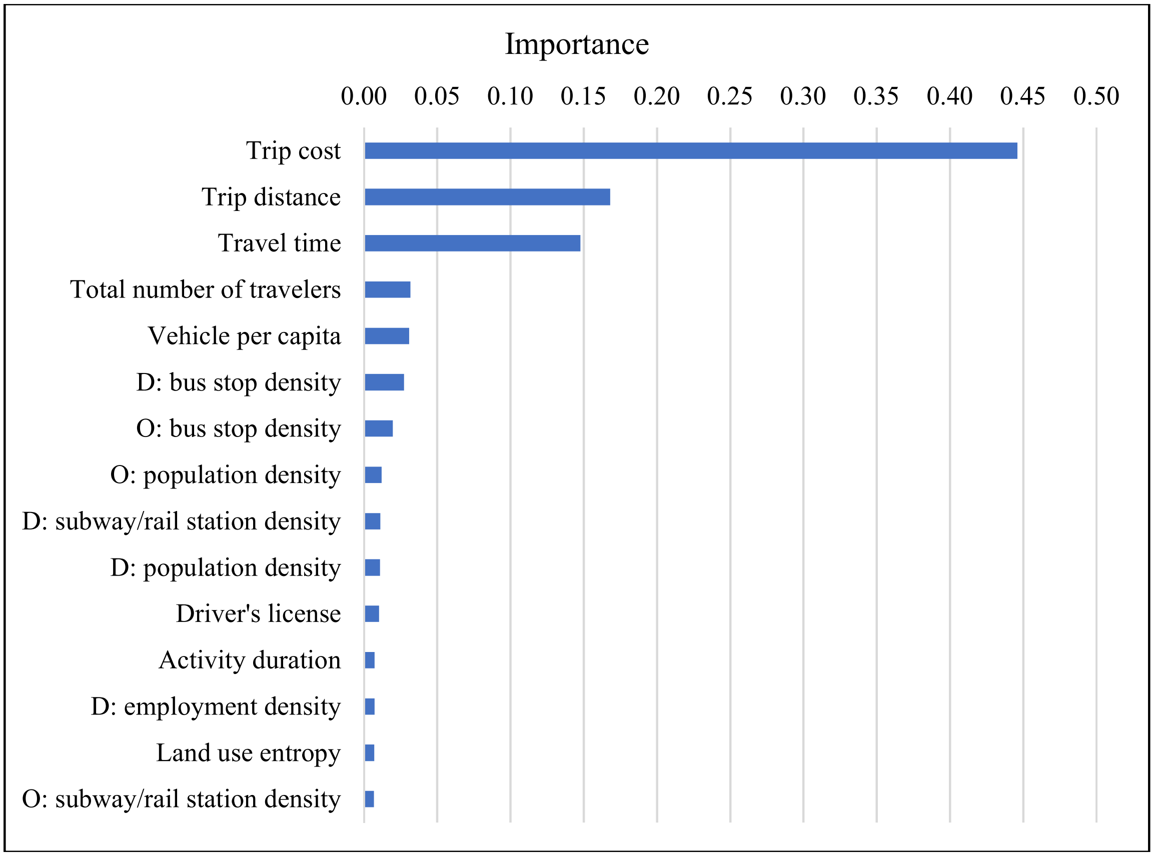


Figure 5.15. Top 15 Important Variables of the XGB Model (Delaware Valley)

5.4.5 IIA Tests of the MNL Models

One of the most important assumption of the MNL model is the "independence of irrelevant alternatives" (IIA) property. In order to examine if the IIA property hold for the developed MNL models, for each developed MNL model, three Nested Logit (NL) models are developed, the structures of which are shown in Figure 5.16. The first NL model combines car and ODRS into a nest, since both modes are automobile based and are often more expensive than others. The second model combines transit and ODRS into a nest, since ODRS shares similarities with transit, such as both are publicly available and are often more accessible in denser areas. There are other similarities as suggested by previous analysis suggests that disabled travelers rely more on ODRS and transit than other travel modes. The third NL model combines walking and biking into a nest, since both modes are non-motorized, more active, and have lower costs.

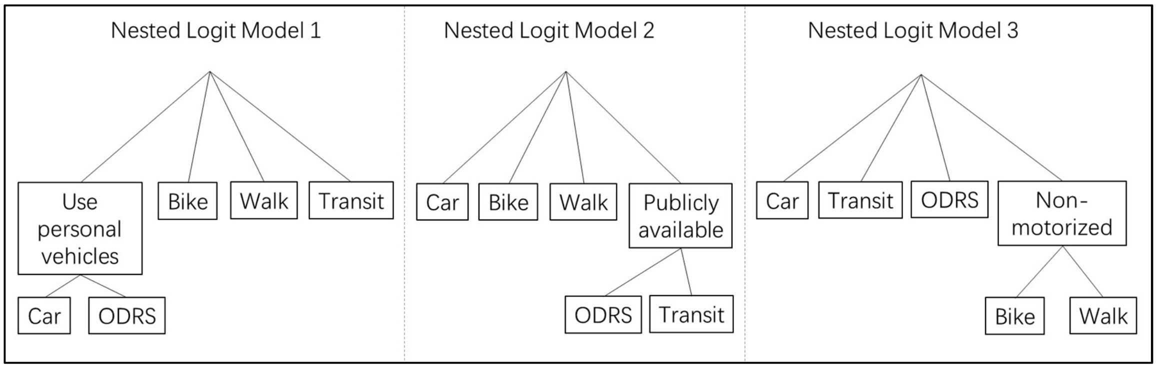


Figure 5.16. Nested Logit Models

The development of MNL models is based on the important assumption of IIA that assumes the independence of irrelevant alternatives (IIA), and in this case, travel modes. In order to test whether IIA holds for the developed models, nested logit (NL) models are developed to examine whether the nest parameter is close to one and whether the nested logit models have significantly better goodness of fit compared to the MNL models. For each of the three hypothesized structures shown in Figure 5.16, a NL model is developed and its result is compared to the MNL model. The results of the T-values of the nest parameter (θ) and Chi-squared tests that compares the goodness of fit of the NL model with the MNL model are summarized in Table 5.20. The nest parameter (θ) of the NL models is an indicator whether the options in the proposed nest have significant similarities. The Chi-square tests examine whether the null hypothesis that the NL model is not statistically better than the MNL model can be rejected at certain confidence level. As the results in Table 5.20 shows, for most of the proposed NL models, the null hypothesis cannot be rejected, indicating that the IIA assumption holds for those corresponding MNL models. This may result from how the developed MNL models are specified, as most of the independent variables included in the MNL models are set as alternative-specific variables. Existing theory has suggested that setting independent variables as alternative-specific rather than generic is often found to be effective for avoiding violation of IIA in MNL models (Mokhtarian, 2016).

Table 5.20. Summary of Nested Logit Models' Results

		NL Model 1 **Combine Car & ODRS**		**NL Model 2** **Combine Transit & ODRS**		**NL Model 3** **Combine Biking & Walking**	
		Value	**Interpretation**	**Value**	**Interpretation**	**Value**	**Interpretation**
NHTS	T-value of the nest parameter	0.38	Cannot reject the null	0.94	Cannot reject the null	0.55	Cannot reject the null
	Chi-squared test	0.00		0.72		0.78	
New York Metro	T-value of the nest parameter	1.01	Cannot reject the null	15.70	**Can reject the null at 99% confidence interval**	0.84	Cannot reject the null
	Chi-squared test	0.02		177.40		0.20	
Puget Sound Region	T-value of the nest parameter	0.64	Cannot reject the null	1.17	Cannot reject the null	0.38	Cannot reject the null
	Chi-squared test	0.57		1.89		0.32	
Delaware Valley Region	T-value of the nest parameter	1.91	Cannot reject the null	4.87	**Can reject the null at 99% confidence interval**	0.90	Cannot reject the null
	Chi-squared test	2.20		21.43		0.43	
* The null hypothesis assumes that the goodness of fit of the NL model is not significantly better than that of the MNL model							
The Chi-squared test compares the log-likelihood of the NL model with the MNL model							

Nevertheless, for the developed NL models using the second nest structure for both the New York region and the Delaware Valley region, the null hypothesis can be rejected at the 99% confidence interval. The two models both use the second proposed nest structure that nests transit and ODRS together. This indicates that for both the New York region and the Delaware Valley region, there are some similarities between transit and ODRS that the MNL models do not account for. The result is very interesting and it further reveals that ODRS plays a similar role as transit under certain circumstances while it may be a challenge to reduce the conflicts between the two modes.

5.5 Conclusions

People's travel mode choices intertwine with many different factors and may have noticeable changes as technological advances and new travel modes and new data sources are available. Modeling travel mode choice is a critical step in travel demand forecasting and may also face challenges and opportunities as new travel modes and data sources become available. The analysis is among the limited number of studies that explore machine learning models' application in travel mode choice modeling and has included a relatively comprehensive list of independent variables that are ready for practical use.

The MNL model and the two machine learning models show high levels of accuracy of predicting the travel mode choices using the four different datasets, but the performance of the two machine learning models surpasses that of the MNL model. The original datasets are extremely unbalanced as ODRS only accounts for about 0.3% in the national data, 0.8% in the New York region and about 0.2% in the Puget Sound and the Delaware Valley region. All three models perform poorly in predicting the choice of ODRS in such unbalanced datasets and the prediction error rates always exceed 80%. After sampling the three datasets to make them more balanced, the performance of the models is improved significantly, though the ODRS-specific prediction errors are still the highest compared to other travel modes. After sampling, the overall prediction accuracy and the mode-specific prediction accuracy of the two machine learning models are both better than the MNL model, indicating the significant advantage of using machine learning for predicting travel mode choices.

The main advantage of the MNL model is its ease of interpretation of results and its high consistency between the training and testing errors. The RF model is also very consistent between the training and testing errors, which is because the design of the RF model makes it very good at avoiding the overfitting issue. The MNL model is the only one that allows direct interpretation of the relationship between the dependent variable and the independent variables and is useful for variable selection and deriving policy implications.

Regarding the effort of developing and implementing the models, the machine learning models and the MNL model have different strengths and challenges. The advantage of the machine learning models is that it has very little limitation on the data structure and model specification. For example, travel distance is included in machine learning models and it is found to be an important independent variable in predicting mode choice, but it cannot be included in the MNL model due to the issue of correlation between travel time and distance. Also, though the machine learning models require more effort of parameter tuning, the whole model fitting process requires less attention compared to the MNL model that requires very careful model specification and testing to examine whether the assumptions hold. In contrast, the MNL model can easily avoid the overfitting issue. Especially for a very unbalanced dataset, the XGB model may not suffer from overfitting issue for all choices combined, but may have an overfitting issue when predicting the choice with small shares. Therefore, attention needs to be paid to the overfitting issue when using machine learning for travel mode choice modeling with unbalanced data. Machine learning techniques that can handle choice-specific overfitting control will be very helpful in this situation.

A broad set of socio-demographic, economic, and built environment variables are found to be associated with people's choice of ODRS. The variables are very consistent with traditional understanding about what factors influence people's mode choice. An interesting finding is that both low-income and high-income travelers are more likely to choose ODRS, which further confirms the previous finding that ODRS serves both captive and choice users. Although the factors identified in this analysis are like factors that are traditionally considered to be relevant to people's travel mode choices, it does not mean that the choice of ODRS is only influenced by those factors. The factors included in this analysis is limited by what variables are available from the regional household travel survey data. This reveals an important need of collecting new information and data about ODRS to further understanding about what factors influence its choice. An obvious example is that currently ride-sourcing can only be accessed with smartphones and be paid by credit cards, so these conditions can determine whether ODRS can be used for some travelers. However, the information is not available in conventionally collected survey data and thus cannot be incorporated into mode choice modeling. It is critical to collect more data and more comprehensive survey data that take consideration into new variables and factors that are related to ODRS choices, but are omitted in conventional datasets.

CHAPTER 6. IMPACT OF ODRS ON TRANSPORT ACCESSIBILITY AND EQUITY

On-demand ride service (ODRS) resembles the notion of Mobility as a Service (MaaS), that is the "shift away from personally owned modes of transportation towards mobility solutions that are consumed as a service" (Zielinski, 2016). ODRS has attracted great attention recently as the users of Transportation Network Companies (TNCs) grow quickly and automated vehicle technology is progressing. ODRS has two intrinsic characteristics that make it different from both conventional public and private mobility options: it is demand-based and it does not require the traveler to own the vehicle. These two characteristics also make ODRS publicly available and agile to fit into multimodal travel, meaning that it has potential to elevate the service level of public transportation. By providing an attractive alternative to driving and filling gaps in the public transit network, ODRS can potentially reduce auto use and associated environmental impacts and serve more transport-disadvantaged population (Metcalfe & Warburg, 2012; Rayle et al., 2016; Wang & Ross, 2017). Many questions surrounding ODRS remain unknown or inconclusive, while it is exhibiting impacts on many aspects of our transportation system and on our daily travel behaviors.

Improving accessibility to employment and other urban amenities across all modes is an important component of transportation planning. In the US context, there is a considerable gap in the accessibility level by cars and public transportation and the recent literature has focused more on accessibility across travel modes, related equity issues, and closing the car-transit accessibility gap (Boarnet et al., 2017; Fan, 2012; Golub & Martens, 2014; Grengs, 2010). Nevertheless, in practice it is always a challenge to improve transit accessibility because of multiple reasons, such as limited funding and the difficulty of matching transit supply and demand. The growth of ODRS appears to be a new opportunity for elevating accessibility by transit since it is publicly available and can fill the gaps in transit, however, little is understood about the potential accessibility impact of ODRS.

Research Question 3 investigates the potential improvement of job accessibility because of ODRS, responding to the recent fast growth of ODRS users and the growing concern that ODRS is reducing transit usage. Understanding the potential accessibility benefits of ODRS does not take away the concern that ODRS competes with public transportation, but quantifying and measuring the potential accessibility benefits can help reveal strategies to maximize the synergistic effect between ODRS and fixed-route transit. Two specific questions are addressed in this chapter: 1) what is the potential impact of ODRS on transit accessibility to employment? and 2) how does the impact vary across population groups with different income levels?

6.1 Methodology and Data of Research Question 3

The analysis of Research Question 3 takes the Puget Sound region as a case study, but the developed methodology can be easily transferred to other regions with minor changes and using publicly available data sources. Twelve scenarios are developed to capture possible variations in ODRS levels of service. In each of the twelve scenarios, job accessibility is re-estimated and compared to the base scenario that does not consider availability of ODRS. How the accessibility impact varies across different population groups is examined to derive equity-related implications. The detailed methodology is described below.

6.1.1 Study Area

The study area includes 13 large areas (commonly used definition in the region) in the Puget Sound region (Figure 6.1). The Puget Sound region consists of four counties, but some of the peripheral areas are rural without any transit service currently. Consequently, only the 13 large areas that are mostly urban (according to U.S. Census Bureau's definition) within the Puget Sound region are considered as the study area.

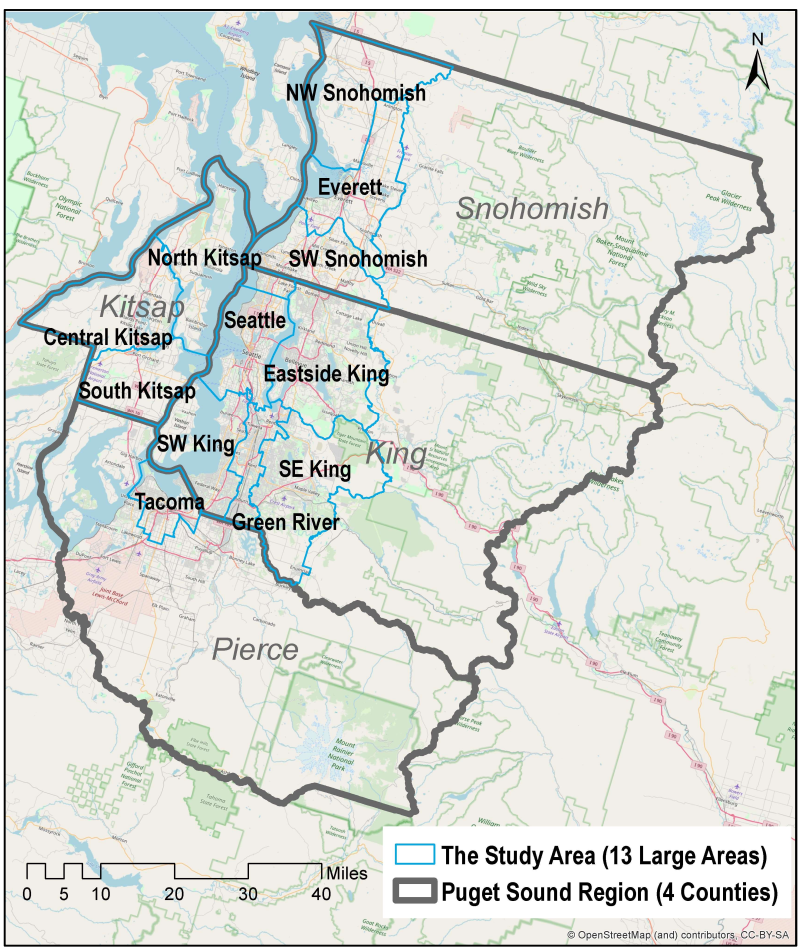


Figure 6.1. The Study Area

6.1.2 Data

The transit network for the study area is composed of three networks: trains and buses operated by Sound Transit, the King County Metro system, the City of Seattle streetcar, the Kitsap transit, the King County marine system, and some community transit. Over 25,000 transit stops and all the transit lines (except community paratransit) in the region are considered in the analysis (Figure 6.2). The transit network dataset is constructed using the General Transit Feed Specification (GTFS) data corresponding to these transit operators and the tool "Add GTFS to a Network Dataset" (Morang, 2014) in ArcGIS. To construct the transit network data in ArcGIS also requires using pedestrian and/or driving road network data. In this dissertation, the pedestrian and driving network data is constructed using the North America Detailed Streets provided by ESRI.

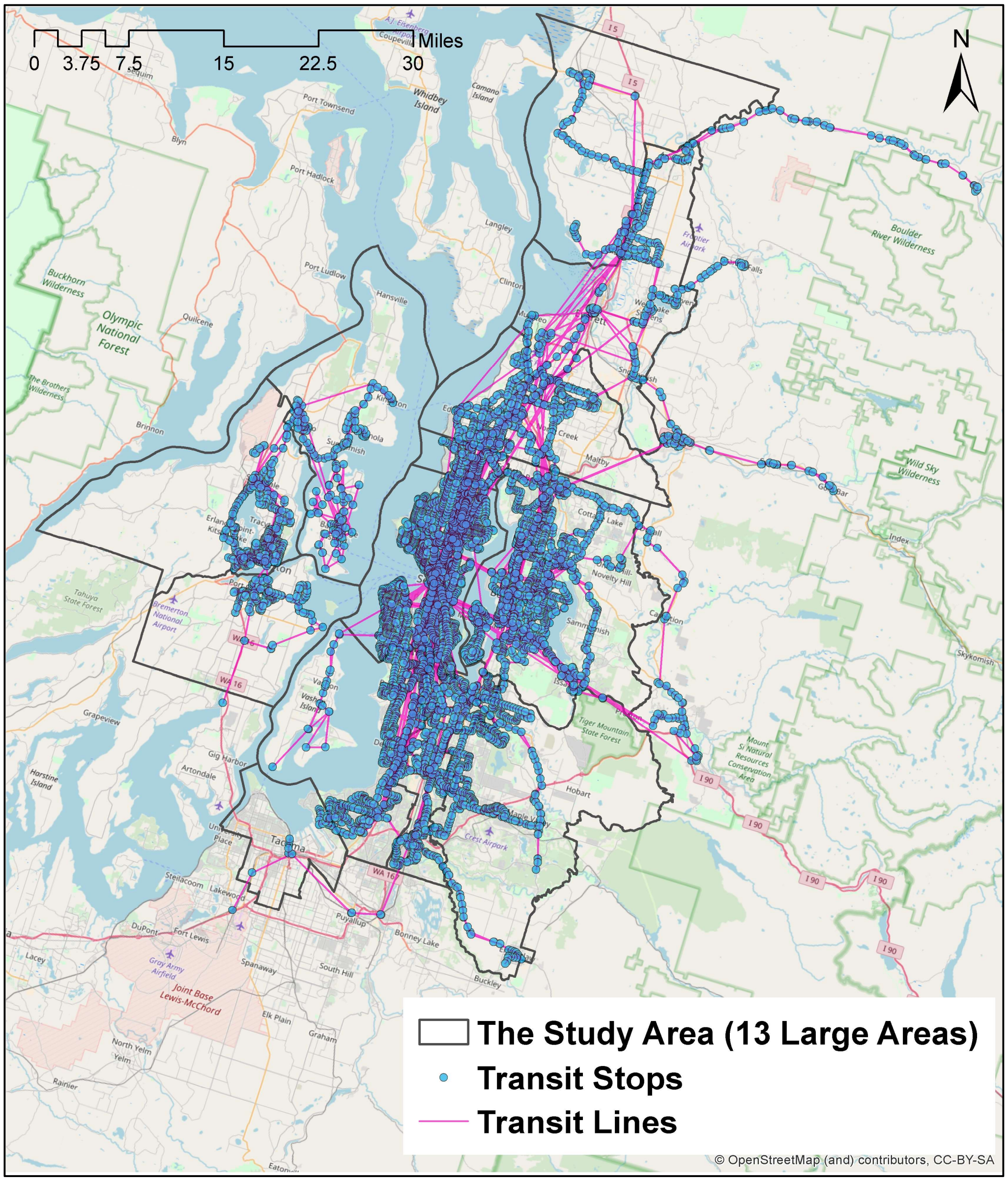


Figure 6.2. Transit Stops and Transit Lines in the Study Area
Note: The transit lines are shown as a network dataset in GIS, which are drawn by directly linking transit stops, so may not show the exact location of the routes.

Resident characteristics are used to segment the population into income subgroups for the equity analysis. Residential data is provided by the Origin-Destination Employment Statistics (LODES) dataset version 7 (LODES 7) from the U.S. Census Bureau's 2015 Longitudinal Employer-Household Dynamics (LEHD) database. The LODES 7 dataset provides Residential Area Characteristics (RAC) for each block group.

Employment data for each census tract is also provided by the LODES 7 dataset, which provides Workplace Area Characteristics (WAC) for each block group. In this research, we follow the definition in LODES 7 to define low-wage, medium-wage, and high-wage jobs as well as low-, medium-, and high-income workers. The definitions are shown in Table 6.1.

Table 6.1. Job and Worker Categories by Wage/Income

Job Categories	Definition	Workers Categories	Definition
Low-wage jobs	Jobs with earnings $1,250/month or less	Low-income workers	Workers with earnings $1,250/month or less
Mid-wage jobs	Jobs with earnings $1,251/month to $3,333/month	Mid-income workers	Workers with earnings $1,251/month to $3,333/month
High-wage jobs	Jobs with earnings greater than $3,333/month	High-income workers	Workers with earnings greater than $3,333/month

6.1.3 Developing the Scenarios

ODRS can potentially affect job accessibility by transit in two direct ways: (1) it can serve the first/last mile of fixed-route public transportation (multimodal travel); and (2) it can directly serve a whole trip from origin to destination (single modal travel). Under the first assumption, ODRS can be used to enlarge the "catchment areas" of transit stops. Under the second assumption, ODRS can be used as paratransit to serve entire trips according to users' request. However, it is not possible to use ODRS to serve all trips, as it may not be available or affordable everywhere. It is also not known exactly what form the future ODRS regime will take and what level of service it can achieve. Therefore, 12 scenarios are developed to reflect different possibilities of ODRS in the study area. Under each of the 12 scenarios, job accessibility is estimated and compared with the base scenario that does not consider availability of ODRS (assuming people are willing to walk up to 0.5 mile to access transit). Equity analysis is developed to compare the difference in ODRS' accessibility impact across low-, medium-, and high- income workers/jobs.

The 12 scenarios are developed based on two parameters, the wait time of ODRS and the travel distance that ODRS can be used for (see Table 6.2). Wait time is used to define the scenarios because it is a main factor determining of the level of service of ODRS. For trips by ODRS, the travel time can be decomposed into in-vehicle travel time and wait time. In-vehicle travel time is simply the same as travel time by driving, so wait time is the major factor that influence the level of service. Just as headway of fixed-route public transportation, the length of wait time of ODRS reflects its level of service: the shorter the wait time, the higher level of service ODRS has.

The second main parameter used to define the 12 scenarios is travel distance that ODRS can be used for, mainly because of two reasons. First, travel distance reflects the cost of using ODRS to some degree, which is currently the major constraint of using ODRS. Considering the potential cost of using ODRS when estimating its impact on accessibility is important. However, the cost of TMCs is still fluctuating, varies by different geographies, and may see significant changes in future if automated vehicles become available. Therefore, using distance as a parameter, rather than cost, provides a more consistent reference for analysis and policy implications. Second, accessibility is intrinsically associated with how many destinations can be reached, so travel distance directly influence accessibility given fixed land use patterns and using different travel distance thresholds captures the impact of ODRS more fundamentally.

Table 6.2. Scenarios of ODRS by Travel Distance and Wait Time

Assumptions	Uniform Wait Time (mean = 6 min)	Demand-based Wait Time (mean =6 min)	Shorter & Uniform Wait Time (mean = 3 min)	Shorter & Demand-based Wait Time (mean = 3 min)
Multimodal travel: 2 miles; Single modal travel: 4 miles.	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Multimodal travel: 1.5 miles; Single modal travel: 3 miles.	Scenario 5	Scenario 6	Scenario 7	Scenario 8
Multimodal travel: 1 mile; Single modal travel: 2 miles.	Scenario 9	Scenario 10	Scenario 11	Scenario 12

6.1.3.1 Wait Time Assumptions

Wait time assumes four forms, corresponding to the four columns in Table 6.2. The first form is a standard, uniform wait time for pick-ups in all block groups in the study area. Although there is no official release of current ODRS wait times from ODRS operation companies, six minutes is a reasonable wait time based on several sources (Fang He & Shen, 2015; Jung, Jayakrishnan, & Park, 2013; Lambert, 2016a). Therefore, Scenarios 1, 5, and 9 assume a uniform six-minute wait time.

The second wait time form is demand-based. Wait time could vary as a function of the density of demand for ODRS service. This is more similar to the reality that ODRS is provided: the denser the area is, the more ODRS service is provided and users may experience shorter wait time of ODRS on average. Accordingly, block groups with greater demand density have shorter wait times than block groups with lower demand density. The combined population and employment density at the block group level is used as a simple representation of the potential demand for ODRS. Block groups are divided into low-demand, mid-demand, and high-demand categories based on their combined population and employment density. Within each category, block groups are randomly assigned wait times between given boundaries, described in Table 6.3. Wait times are randomly assigned to account for real-life variability. The wait times are set to maintain a study-area-wide average wait time of 6 minutes. Demand-based wait times are applied to Scenarios 2, 6, and 10.

Table 6.3 Demand-based Wait Time Definitions

Group	Number of Block Groups	Lower Wait Boundary	Upper Wait Boundary
Short Wait Time	519	1 minute	3 minutes
Medium Wait Time	755	4 minutes	7 minutes
Long Wait Time	896	8 minutes	10 minutes

The third form is short, uniform wait times of 3 minutes, which are applied to Scenarios 3, 7, and 11. This form assumes an increase in the number of vehicles serving ODRS as well as potential improved allocation of ODRS service that allow the ODRS serve customers more quickly.

The final wait time form is a shorter demand-based wait time variation. Similar to the second form, this form also assumes that the wait time of ODRS is associated with the potential demand, but with potential increase in ODRS fleet size or improved service, the wait time may be shortened. The same categorization of block groups is used to group them into low-demand, mid-demand, and high-demand block groups according to the combined population and employment density. Within each category, block groups are randomly assigned wait times between given boundaries, described in Table 6.4. Wait times are randomly assigned to account for real-life variability. The wait times are set to maintain a study-area-wide average wait time of 3 minutes to reflect the potential improvement in level of service of ODRS, compared to the second form. This shorter demand-based wait times are applied to Scenarios 4, 8, and 12.

Table 6.4. Shorter and Demand-based Wait Time Definitions

Group	Number of Block Groups	Lower Wait Boundary	Upper Wait Boundary
Short Wait Time	519	1 minute	1.5 minutes
Medium Wait Time	755	1.5 minutes	3.5 minutes
Long Wait Time	896	6.5 minutes	6 minutes

6.5.1.1 Travel Distance Assumptions

Three assumptions of distance thresholds are used to approximate monetary cost and other constraints of using ODRS, corresponding to the three rows in Table 6.2. The multimodal travel distance threshold provides the number of miles that a traveler is willing to take ODRS to access transit, while the single modal distance threshold provides the distance that a traveler is willing to take ODRS to access a destination (without transit). These two thresholds correspond to the two ways that ODRS can influence accessibility: (1) by serving the first/last mile of transit (multimodal travel) and (2) by serving an entire trip from origin to destination (single modal travel). The single modal distance threshold is always two times of the multimodal distance thresholds, assuming that if a traveler is willing to use ODRS for certain travel distance at both ends of a transit trip, the traveler would be willing to simply use ODRS for twice the length in a single modal travel.

For Scenarios 1 through 4, it is assumed that ODRS can be used for a trip up to 2 miles when it is used as an access/egress mode to transit, while it can be used for a trip up to 4 miles when it is serving a whole trip from origin to destination. Similarly, for Scenarios 5 through 8, it is assumed that the maximum distance of transit-accessing/egressing trip is 1.5 mile while the maximum of a single modal trip is 3 miles. For Scenarios 9 through 12, the maximum distance of an accessing/egressing trip is 1 mile while the maximum for a single modal trip is 2 miles. It is found that more than 95% of the transit-accessing/egressing taxi trips in New York City are between 0.5 to 2 miles (Wang & Ross, 2017), so the length of transit-accessing/egressing trips is capped to 2 miles.

6.5.2 *Measuring Accessibility to Employment*

Accessibility to employment is measured using the index shown by Equation (6.1) which is a hybrid of a cumulative opportunity and gravity measure of accessibility, suggested by the literature (Karner, 2018). When the accessibility index is higher, it means that the residents in a given block group can reach more jobs by transit. The best travel time by transit and/or ODRS, T_{ij}, is calculated for each pair of block groups located in the study area.

$$\boldsymbol{A_i^w = \sum_j E_j^w * e^{-\beta * T_{ij}}} \qquad \textbf{Equation (6.1)}$$

Where

A_i^w = Accessibility at block group i for employed residents (workers) with wage level *w*;

E_j^w = Jobs in block group j with wage level *w*;

T_{ij}= Travel time (minutes) by transit and/or ODRS between block group i and block group *j*

β = Empirically derived impedance term: β = 0.031 if $T_{ij} > 45$; β = 0 otherwise.

The impedance term β is derived using the Puget Sound regional household travel survey data by fitting an exponential decay function against the trip frequency by travel time and is estimated as 0.031. The average travel time by transit in the Puget Sound region is 45 minutes, so it is assumed that the propensity to travel by transit only decreases when the travel time is longer than 45 minutes. This explains why β is set to be 0.031 only when the travel time is longer than 45 minutes in Equation (6.1). This means that all job opportunities within the 45 minutes are weighted equally but those further away have smaller weights. The exponential decay function shrinks fast, meaning that when travel time exceeds 45 minutes, the weight of a job reduces significantly as the travel time increases. Figure 6.3 shows what the value of 10 jobs, 100 jobs, and 1000 jobs become in the accessibility index as the travel time changes from 0 to 200 minutes using Equation (6.1).

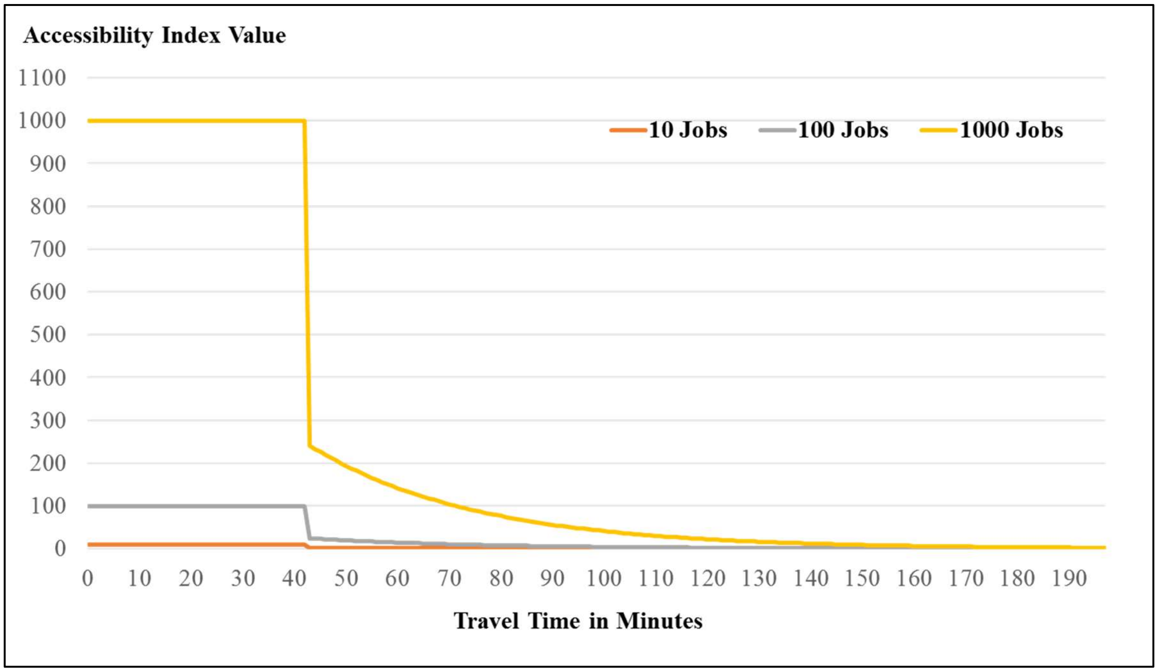


Figure 6.3. Change of the Accessibility Index by Travel Time

Job accessibility in the study area is estimated for every block group in the study area in the base scenario and in each of the 12 scenarios using Equation (6.1). The base scenario assumes that people are willing to walk for up to 0.5 miles to access transit stops and are also only willing to walk for up to 0.5 mile from a transit stop to reach their destinations. In all scenarios, including the base and the 12 scenarios, if the centroids of two block groups are located within 0.5 miles from each other, the two block groups are considered accessible by walking, so adding ODRS to the estimation will not change the travel time between the two block groups. Otherwise, the base scenario is not considering the availability of ODRS, so that in each of the 12 scenarios, travel time to/from transit stops, and the transit stops serving a block group are changed because of ODRS.

To also examine the potential impact of ODRS on transit service equity, wage-level-specific accessibility are estimated in all 12 scenarios. With the estimated best travel time between block group pairs, the accessibility indexes in Equation (6.1) can be adjusted to calculate accessibility to employment of four categories: (1) all jobs, (2) low-wage jobs, (3) mid-wage jobs, and (4) high-wage jobs. The wage categories are defined in Table 6.1. Understanding how accessibility to employment of different wage categories change in each of the 12 scenarios compared to the base scenario has important implications for understanding the equity impact of ODRS.

6.5.3 Travel Time Estimation and Implementation

The impact of ODRS on job accessibility is mainly measured by estimating the change in shortest travel time between block groups and the resulted change in number of jobs that can be reached from a block group. Therefore, T_{ij} in Equation (6.1) is the major parameter that needs to be re-estimated in each of the 12 scenarios. Estimating T_{ij} follows the four major steps (see Figure 6.4).

Step 1: Service areas of all the transit stops are developed differently in different scenarios to identify the transit stops serving each block group in the 12 scenarios. In the base scenario, the service areas are developed as 0.5-mile (pedestrian network distance) area from a transit stop. In contrast, in Scenario 1 through 4, the service areas are developed as 2-mile (driving network distance) areas from transit stops, because in these scenarios, it is assumed that people can take ODRS for up to 2 miles to access transit. Similarly, in Scenarios 5 through 8, the service areas are developed as 1.5-mile area around transit stops, and in Scenarios 9 through 12, the service areas are developed as 1-mile area around transit stops.

Step 2: Transit-stop-pairwise travel times are estimated using the ArcGIS Network Analyst. The GTFS data allows estimation of transit travel time between any origin-destination pair for specified trip departure time. Ten departure time points between 7am to 9 am on a normal Tuesday in 2016 are randomly generated and all transit-stop-pairwise travel times are estimated using the constructed transit network dataset corresponding to the ten departure times. The average of the ten sets of transit-stop-pairwise travel times is used as the final transit-stop-pairwise travel times, to avoid any bias resulting from different departure times. Transit-stop-pairwise travel times are fixed in different scenarios, assuming fixed level of service of the transit network.

Step 3: To implement the assumptions in the 12 scenarios, a searching script is used to identify the shortest travel time between block groups, based on the previous two steps. More specifically, for a given block-group pair, all possible pairs of transit stops that serve this block-group pairs are identified. For the origin and destination block groups, the assumed wait time of ODRS is pre-assigned. Therefore, for the given block-group pair, all possible multimodal travel times, which is the sum of the transit-stop-pairwise travel time and travel time of ODRS (sum of ODRS wait time and in-vehicle travel time), can be estimated.

Step 4: For a given scenario, a searching script is then used to compare two types of travel times for a given block-group pair and find the minimum between the two: (1) minimum multimodal travel time (by transit and ODRS) identified in Step 3; and (2) single modal travel time by ODRS. The second type of travel times is computed as the sum of ODRS wait time at the origin block group and the in-vehicle travel time by ODRS from the origin block group to the destination block group. The in-vehicle travel time of ODRS is estimated as the free-flow travel time by driving for the network distance between the block-group pairs. In Scenarios 1 through 4, for block group pairs that are closer than 4 miles from each other, it is assumed that travelers can simply take ODRS without transit to travel between the block group pairs; In Scenarios 5 through 8, travelers can take ODRS to travel between block groups that are closer than 3 miles from each other; In Scenarios 9 through 12, travelers can take ODRS to travel between block groups that are closer than 2 miles from each other. Consequently, for a given block group pair in a specified scenario, after comparing the shortest multimodal travel time and the single modal travel time (if the distance between the block group pair is within the single modal distance threshold in the scenario), the minimum of the two is then used as the travel time T_{ij} for the block-group pair in Equation (6.1).

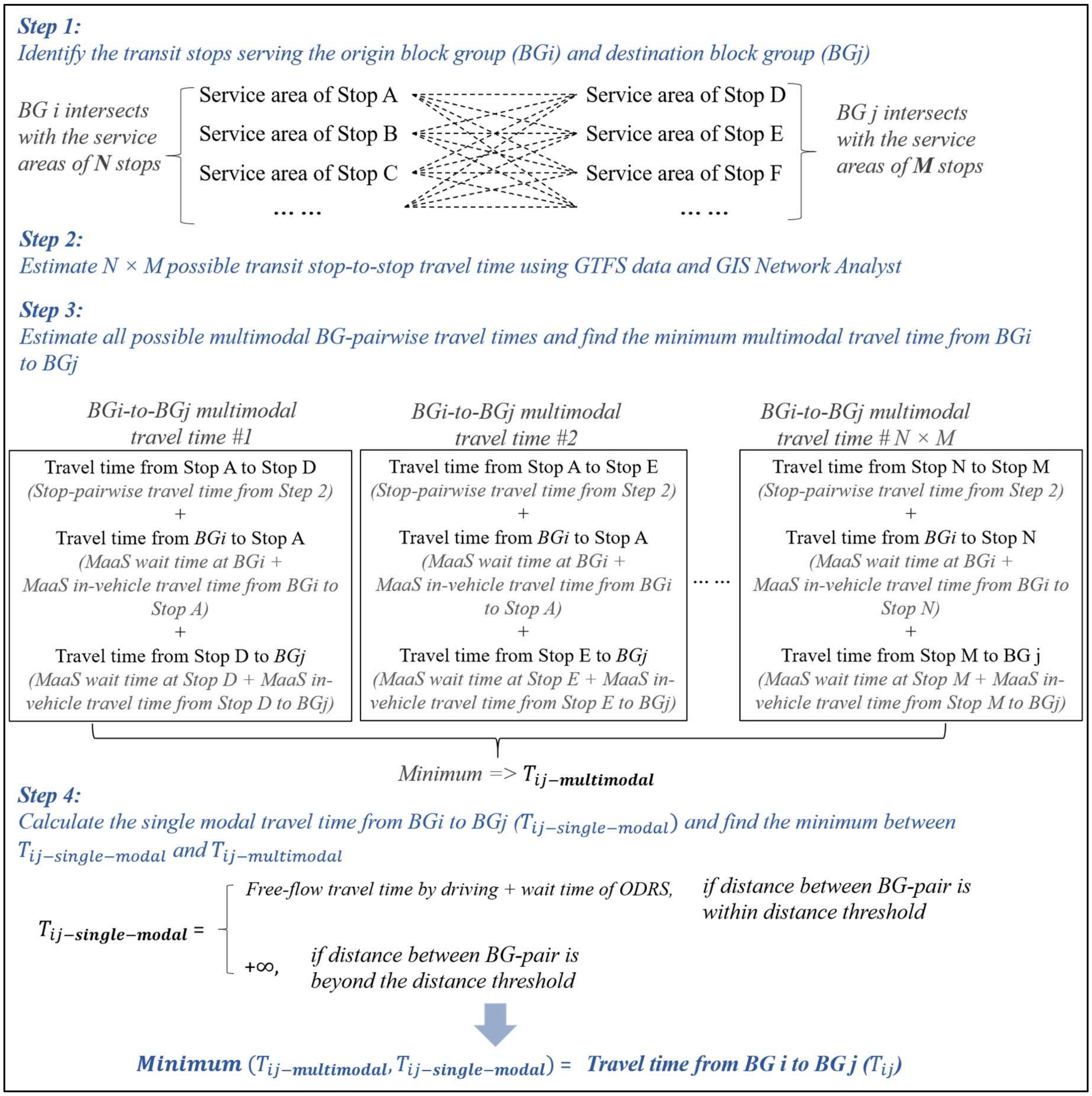


Figure 6.4. Methodology of Estimating Travel Time between a Block Group Pair

There are 2,170 block groups in the study area and 4,708,900 block group pairs in total. According to the Puget Sound regional household travel survey data, more than 98% of trips made by transit are less than 200 minutes, so the cut-off point of travel time is set as 200 minutes, meaning that if the travel time by transit between two block groups are longer than 200 minutes, the two block groups are considered not accessible to each other at all and will not be considered in accessibility estimation using Equation (6.1).

6.6 Travel Time by Transit in the Puget Sound Region

With assumptions in the scenarios changing, the number of block group pairs have travel time less than 200 minutes also changes (see Table 6.5). Since 200 minutes the cutoff point of travel time for a block group pair to be considered in the accessibility analysis, it is important to benchmarking the number of block group pairs with travel time less than 200 minutes to see the changes because of ODRS. In the base scenario, there are 363,644 block group pairs that have travel time by transit less than 200 minutes, and that number increases to 631,641 for Scenario 9 – 12, and 802,267 for Scenarios 5 – 8, and 955,655 for Scenarios 1 – 4. This reflects that the availability of ODRS have significant positive impact on accessibility to employment and of course when ODRS can be used for longer distance, the impact is more significant. As the statistics in Table 6.5 shows, using ODRS has significant influence on the shortest travel time between block groups. even when the total number of block group pairs considered in the analysis increase significantly in the 12 scenarios compared to the base scenario, the average travel times in those scenarios are shorter than the average in the base scenario, indicating that ODRS significantly shortens travel time regionwide. The average travel time of the 1st quantile of block group pairs drops most significantly, which is consistent with our intuition, as ODRS can most significantly improve travel time for shorter distance by providing point-to-point service between two block groups that are close to each other.

Table 6.5. Descriptive Statistics of Travel Time between Block Group Pairs

Scenarios			No. of Block Group Pairs with Travel Time < 200 min	Min (mins)	1st Quantile (mins)	Median (mins)	Mean (mins)	3rd Quantile (mins)	Max (mins)
Base Scenario (S Base)			363,644	0	53	79	72	95	110
S 1	Longest travel distance threshold	Uniform WT	955,655	0	37	71	66	96	120
S 2		Demand-based WT	955,655	0	36	70	65	96	128
S 3		Shorter & Uniform WT	955,655	0	31	66	61	91	114
S 4		Shorter & Demand-based WT	955,655	0	31	65	61	91	118
S 5	Moderate travel distance threshold	Uniform WT	802,267	0	42	74	68	97	118
S 6		Demand-based WT	802,267	0	41	74	67	96	126
S 7		Shorter & Uniform WT	802,267	0	37	69	63	92	112
S 8		Shorter & Demand-based WT	802,267	0	37	69	63	91	116
S 9	Shortest travel distance threshold	Uniform WT	631,641	0	47	77	70	97	116
S 10		Demand-based WT	631,641	0	46	77	69	96	124
S 11		Shorter & Uniform WT	631,641	0	43	73	66	93	110
S 12		Shorter & Demand-based WT	631,641	0	43	73	66	92	114

Note: S 1 – 12 stands for Scenario 1 through 12; “WT” stands for “wait time”. See more details about the 12 scenarios in Section 6.1.3

6.7 Current Accessibility to Employment

The study area has over 1.6 million jobs and about 19.2% of the jobs are low-wage, 27.8% are mid-wage jobs, and 53.0% are high-wage (see Table 6.6). The spatial distribution of low-, mid-, and high- wage jobs are similar, with higher concentration in the metropolitan centers such as Everett, City of Seattle, and Tacoma, and less jobs in the peripheral area. Following the same definition of low-, mid-, high-wage jobs, the workers can also be classified into low-, mid-, and high-income. There are over 1.4 million workers residing in the study area, and about 19% of them are low-income, 28.2% are middle-income, and 52.8% are high-income workers.

Table 6.6. Employment and Workers by Wage Categories in the Study Area

Total Jobs		Low-wage Jobs		Mid-wage Jobs		High-wage Jobs	
No.	%	No.	%	No.	%	No.	%
1,666,628	100.0%	320,059	19.2%	463,536	27.8%	883,033	53.0%
Total Workers		Low-income Workers		Mid-income Workers		High-income Workers	
No.	%	No.	%	No.	%	No.	%
1,436,141	100.0%	272,475	19.0%	405,094	28.2%	758,572	52.8%

There is great variation in current accessibility to employment in the study area. Accessibility to all employment, low-wage, mid-wage, and high-wage employment are mapped in Figure 6.5. For all types of jobs, the areas from Everett to Seattle and to Tacoma have the highest concentration of employment, while some peripheral areas have far less employment and much lower employment density. The descriptive statistics shown in Table 6.7 also reveal the significant variation in accessibility at the block group level. The median accessibility to all employment is 21,613, while the 1st quantile is only 3,614 and the 3rd quantile is 73,165. Some block groups have extremely high accessibility, especially to high-wage jobs, as the maximum of accessibility to all employment is 369,964, while 216,520 of it is for high-wage jobs. The areas with highest accessibility levels are the areas in the center of the region, mostly the areas from Everett to Seattle to Tacoma, which have seen fast growth of high-wage jobs in the information and technology industry in recent years.

Table 6.7. Descriptive Statistics of Accessibility in the Base Scenario

	Accessibility in Base Scenario			
	All Jobs	Low-wage Jobs	Mid-wage Jobs	High-wage Jobs
Min	0	0	0	0
1st Quantile	3,614	775	1,289	1,232
Median	21,613	4,350	6,727	8,990
3rd Quantile	73,165	14,414	20,776	36,694
Max	369,964	77,091	81,677	216,520
Average	52,606	10,942	13,854	27,810

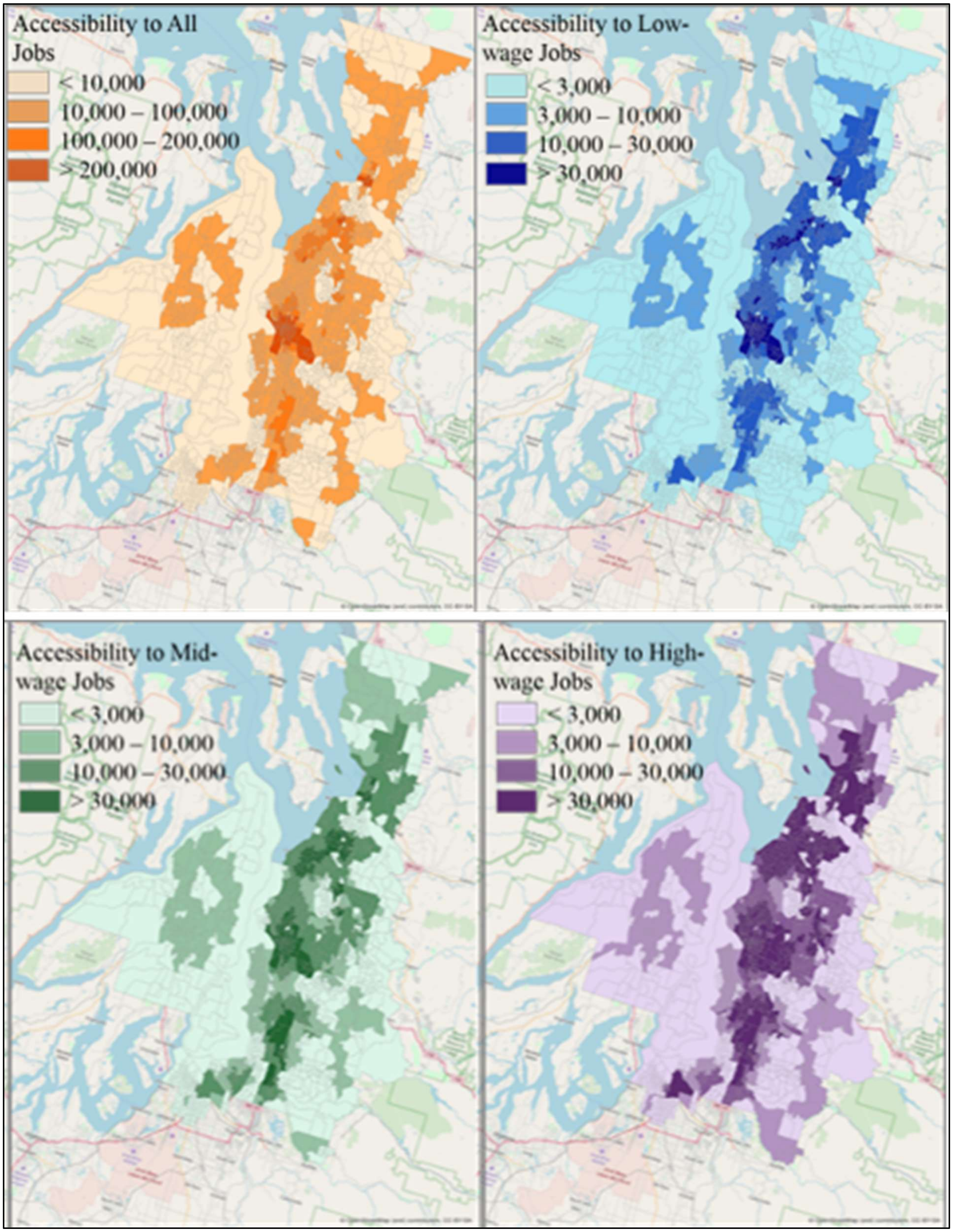


Figure 6.5. Current Accessibility to Jobs of Different Wage Levels

6.8 Impact of ODRS on Accessibility and Equity

Job accessibility by wage categories are estimated in each of the 12 scenarios that considers the availability of ODRS. Overall, accessibility to employment is significantly augmented in the 12 scenarios, as shown in Figure 6.6. Percent change of job accessibility in the twelve scenarios and some descriptive statistics are summarized in Table 6.8 and in Figure 6.7. In Scenarios 1 to 4, which assume that ODRS can be used for up to 2 miles around transit stops and up to 4 miles for single modal trips, the accessibility have the most significant increase. On average, the block group level accessibility in Scenarios 1 and 2 increases by about 220% compared to the base scenario, while the average increases by about 250% when the average wait time of ODRS drops to 3 minutes in Scenarios 3 and 4. For Scenarios 5 to 8, it is assumed that ODRS can be used for up to 1.5 miles around transit stops and 3 miles for single modal trips. The average accessibility increase is about 150% when the average wait time of ODRS is 6 minutes and 175% when the average wait time is shortened to 3 minutes. For Scenarios 9 to 12, it is assumed that ODRS can be used for only up to 1 mile around transit stops and for 2 miles for single modal trips. The average accessibility increase is about 86% when the average wait time of ODRS is 6 minutes and about 101% when the average wait time is shortened to 3 minutes.

The accessibility changes by block group quantiles are also summarized in Table 6.8 and an interesting finding is revealed. As the table shows, in all 12 scenarios, the block groups in the 1st quantile has the most significant accessibility increase because of ODRS and the percent increase is far more than that of the higher quantiles. Ranging from Scenario 1 to 12, the averages of percent accessibility increase of the 1st quantile of block groups range from 476% to 1,447%, while the average increase only ranges from 86% to

254% for all quantiles combined. This is not hard to understand, as block groups with lower accessibility currently will see higher percentage growth when nearby employment become accessible because of ODRS. Nevertheless, the accessibility increase is also very substantial in terms of its absolute quantity, as in Scenarios 9 to 12, the 1st quantile of accessibility to all employment is more than 20 thousand, while in the base scenario, the 1st quantile is only 3,614. This implies that enlarging the catchment areas of transit stops by ODRS and using ODRS for short-distance trips can significantly improve accessibility levels region-wide, but the increase is especially substantial in areas with low current accessibility.

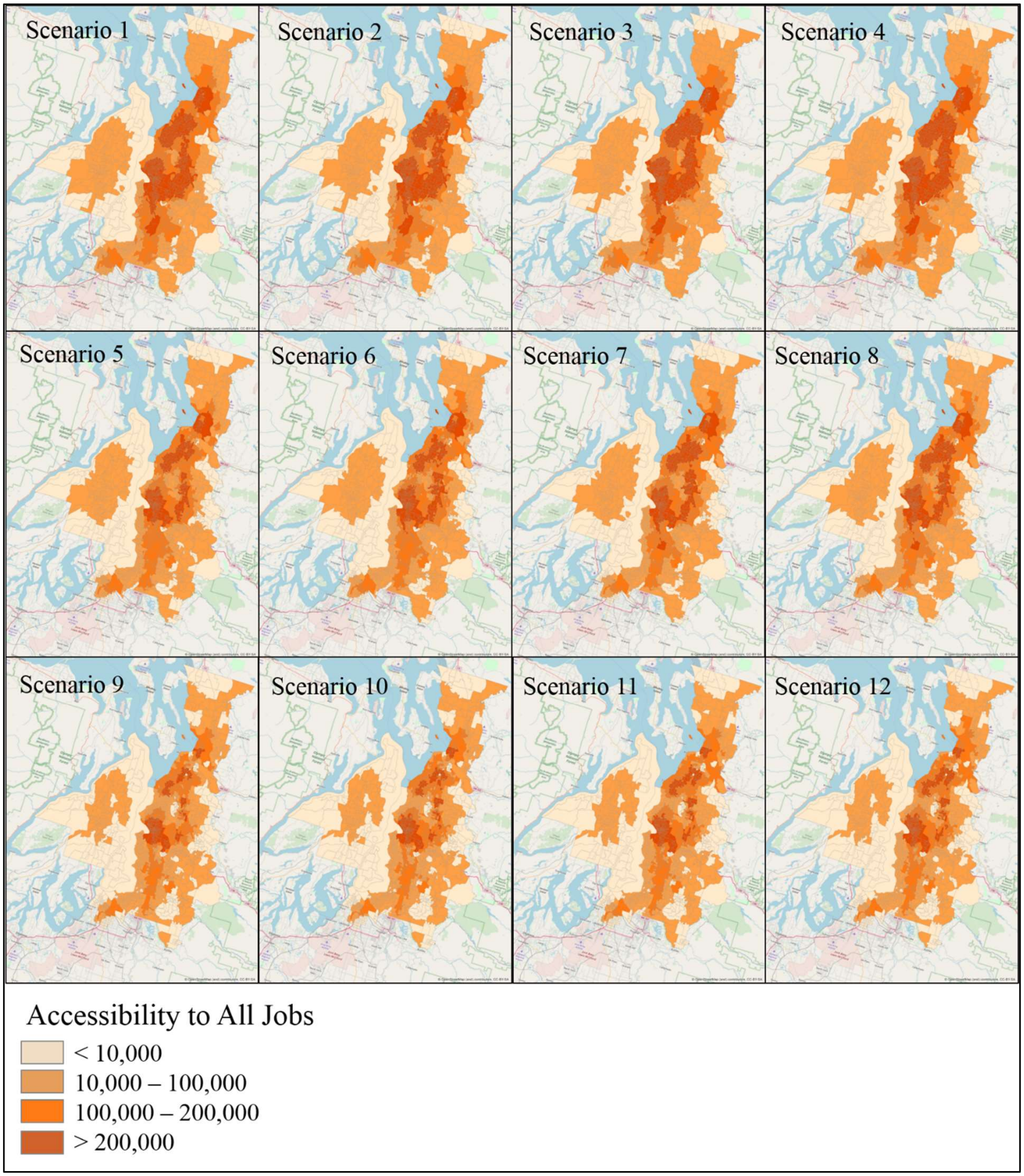


Figure 6.6. Block-Group Level Accessibility to All Jobs in the Twelve Scenarios

Table 6.8. Change of Accessibility to All Employment in the 12 Scenarios

Overall Accessibility Change

	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	25	n.a.	25	n.a.	25	n.a.	25	n.a.
1st Quantile	54,584	1410%	54,584	1410%	55,910	1447%	55,815	1444%
Median	135,264	526%	132,140	511%	151,575	601%	150,689	597%
3rd Quantile	233,489	219%	237,448	225%	260,744	256%	264,855	262%
Max	557,560	51%	638,059	72%	600,408	62%	658,356	78%
Average	169,238	222%	169,797	223%	183,807	249%	186,013	254%
	Scenario 5		Scenario 6		Scenario 7		Scenario 8	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	10	n.a.	10	n.a.	10	n.a.	10	n.a.
1st Quantile	36,126	900%	36,091	899%	37,447	936%	37,457	936%
Median	99,600	361%	100,844	367%	109,337	406%	111,876	418%
3rd Quantile	172,896	136%	178,803	144%	191,531	162%	195,616	167%
Max	505,263	37%	537,972	45%	547,776	48%	563,076	52%
Average	132,179	151%	134,053	155%	143,933	174%	145,965	177%
	Scenario 9		Scenario 10		Scenario 11		Scenario 12	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	10	n.a.	10	n.a.	10	n.a.	10	n.a.
1st Quantile	20,835	476%	20,991	481%	21,743	502%	21,800	503%
Median	66,846	209%	67,732	213%	73,328	239%	73,938	242%
3rd Quantile	129,065	76%	130,482	78%	138,645	89%	141,463	93%
Max	446,991	21%	463,205	25%	461,700	25%	486,894	32%
Average	97,841	86%	100,323	91%	105,542	101%	108,585	106%

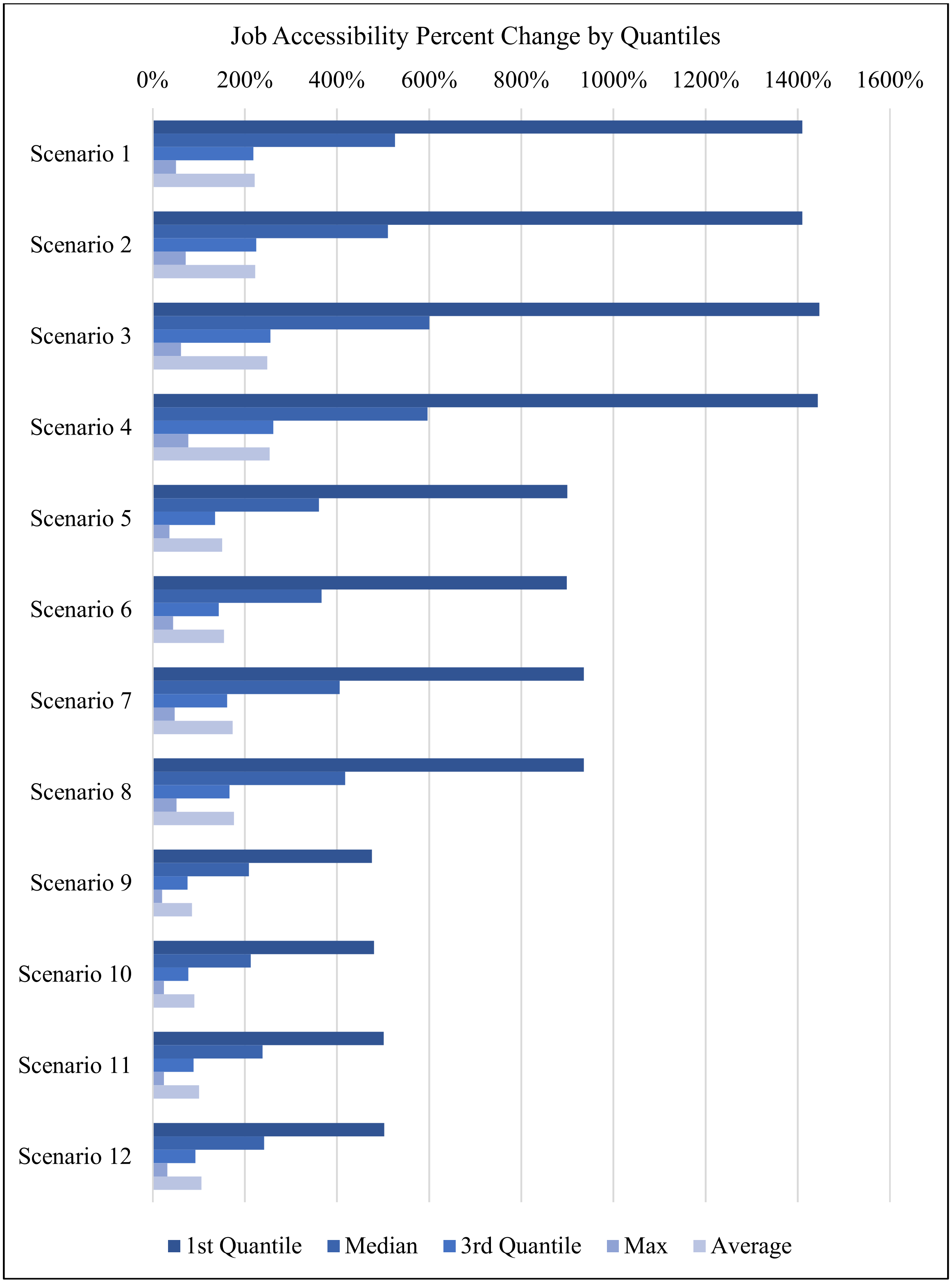


Figure 6.7. Block-Group Job Accessibility Percent Change by Quantiles

In addition to understanding the impact of ODRS on overall accessibility, it is also important to understand the potential change to accessibility to low-wage employment, simply because transit-dependent population are more likely to be low-income population. The spatial distributions of accessibility to low-wage jobs in all the 12 scenarios are mapped in Figure 6.8. Generally, the growth of accessibility to low-wage jobs in different scenarios follows a similar pattern of the overall accessibility change. Descriptive statistics are shown in Table 6.9 and Figure 6.9 provides a quick comparison of the percent change of accessibility to low-wage jobs in different scenarios. Though the region has smaller number of low-wage jobs compared to mid- and high- wage jobs, the percent growths of accessibility to low-wage jobs in the 12 scenarios are similar compared to the percent growths of accessibility to all types of jobs. Even in scenarios that assume ODRS wait time varies according to potential demand (low-income areas have longer wait time of ODRS), such as in Scenarios 2, 4, 6, 8, 10, 12, the increase of accessibility to low-wage jobs is also very substantial. The significant accessibility improvement to low-wage jobs indicates that the accessibility benefit of ODRS is evenly distributed across wage levels.

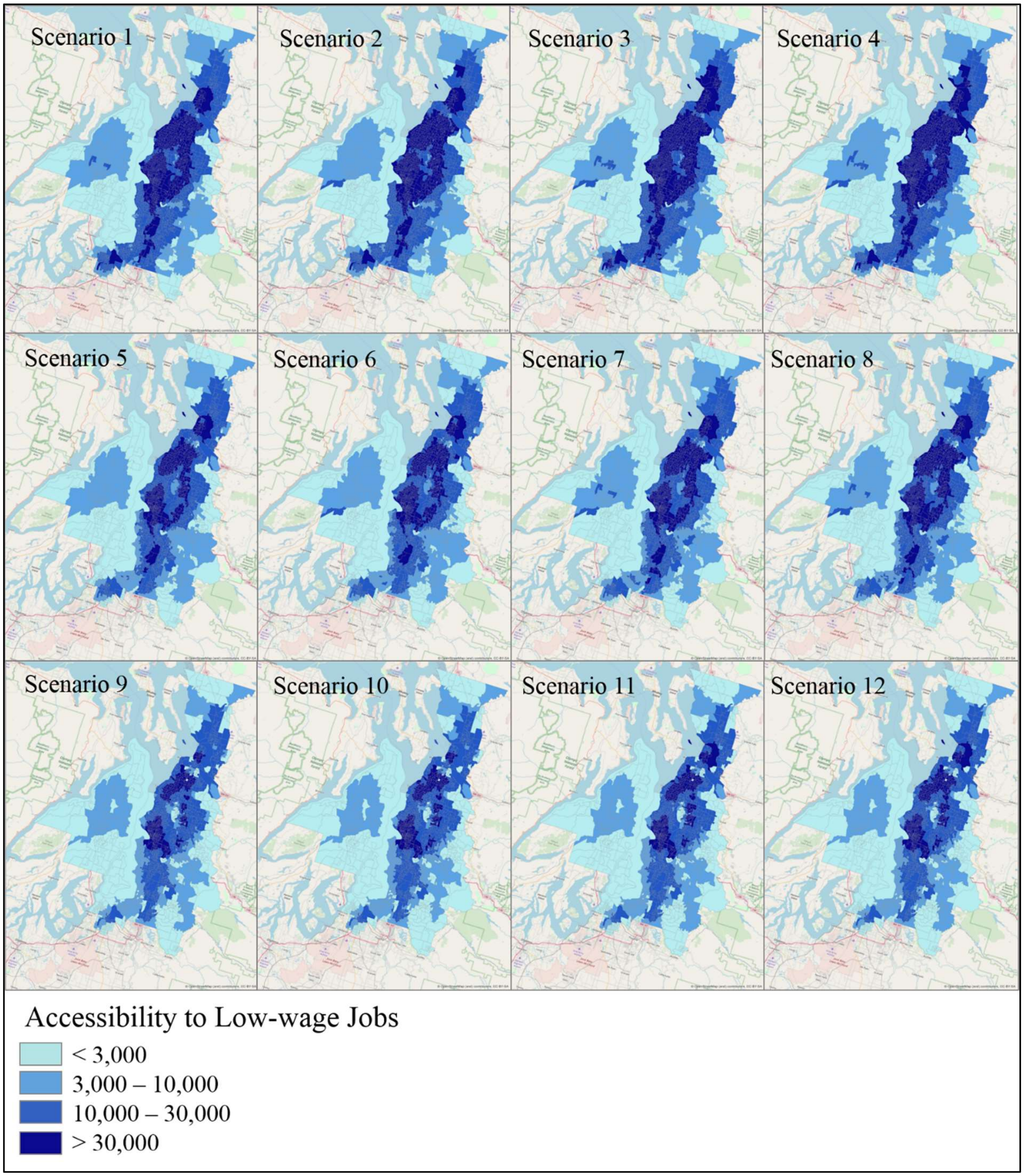


Figure 6.8. Accessibility to Low-Income Jobs in the Twelve Scenarios

Table 6.9. Change of Accessibility to Low-wage Jobs in the Twelve Scenarios

Low-wage Job Accessibility Change								
	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	15	n.a.	15	n.a.	15	n.a.	15	n.a.
1st Quantile	11,033	1324%	10,985	1317%	11,412	1373%	11,408	1372%
Median	25,203	479%	23,676	444%	27,548	533%	27,108	523%
3rd Quantile	42,005	191%	42,922	198%	46,586	223%	47,673	231%
Max	113,840	48%	116,732	51%	121,664	58%	121,558	58%
Average	32,572	198%	32,741	199%	35,304	223%	35,738	227%
	Scenario 5		Scenario 6		Scenario 7		Scenario 8	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	3	n.a.	3	n.a.	3	n.a.	3	n.a.
1st Quantile	7,716	896%	7,731	898%	8,337	976%	8,216	960%
Median	18,643	329%	18,397	323%	20,251	366%	20,205	364%
3rd Quantile	31,651	120%	32,556	126%	34,974	143%	36,105	150%
Max	104,118	35%	108,816	41%	111,506	45%	114,300	48%
Average	25,695	135%	26,103	139%	27,933	155%	28,349	159%
	Scenario 9		Scenario 10		Scenario 11		Scenario 12	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	3	n.a.	3	n.a.	3	n.a.	3	n.a.
1st Quantile	4,755	513%	4,721	509%	4,960	540%	4,970	541%
Median	11,830	172%	11,864	173%	12,965	198%	12,941	197%
3rd Quantile	24,094	67%	24,974	73%	26,459	84%	26,615	85%
Max	91,604	19%	99,020	28%	98,790	28%	101,287	31%
Average	19,123	75%	19,603	79%	20,578	88%	21,201	94%

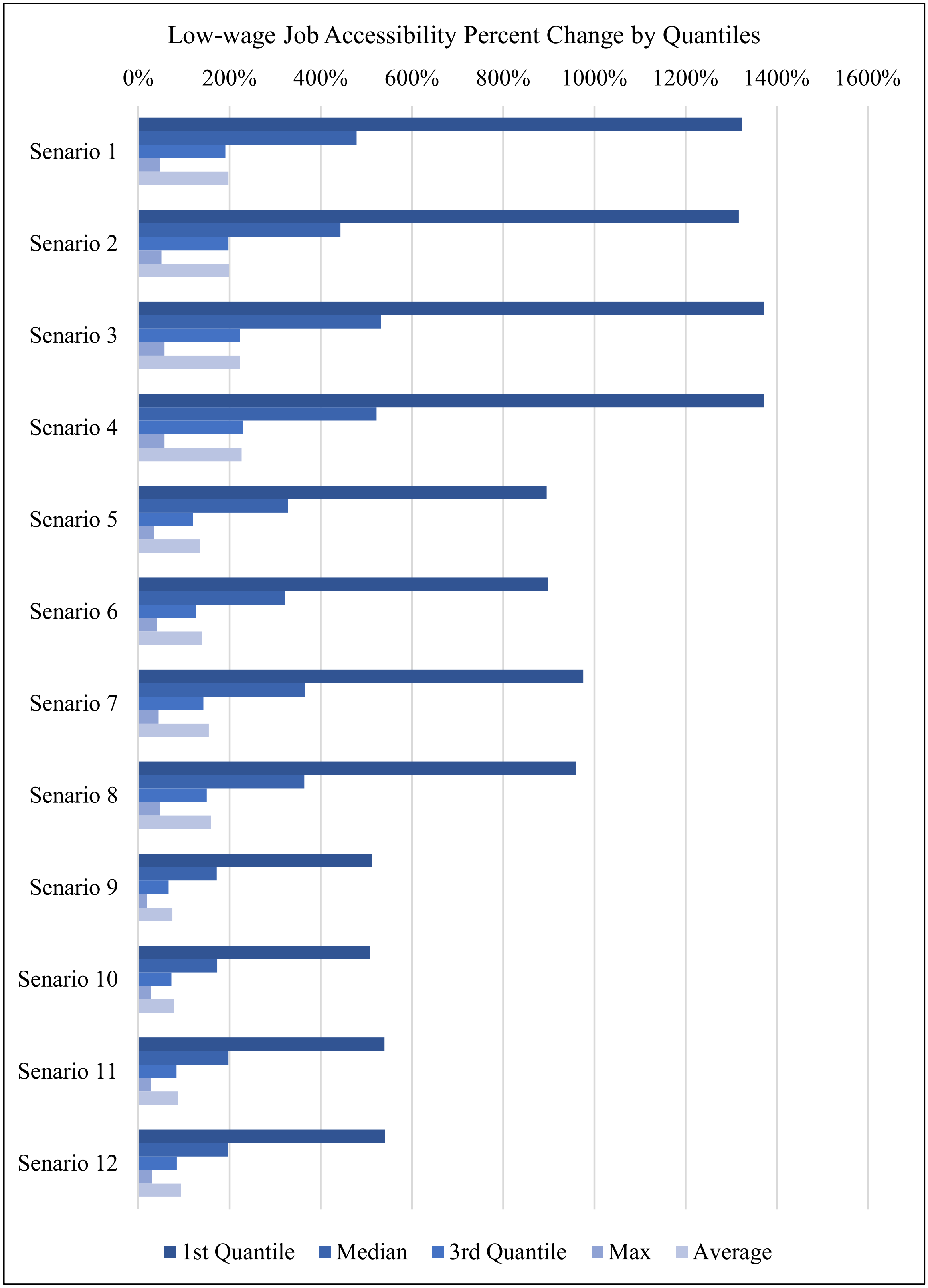


Figure 6.9. Low-wage Job Accessibility Percent Change by Quantiles

The spatial mismatch between jobs and housing has been studied for a long time, so in addition to examining the potential change in accessibility to employment, it is also necessary to compare the spatial distributions of job accessibility to where workers are. Job-to-worker ratio is a straightforward metric that reflects how many jobs are accessible per worker. The job-to-worker ratio of a block group is simply calculated as the number of jobs (of certain category) that are accessible from a block group divided by the number of workers (of the same category) residing in the block group. In the base scenario, the average job-to-worker ratio for the study area is 103, indicating that 103 jobs are accessible by transit for every worker on average. The average low-wage job to low-income worker ratio is 111; mid-wage job to mid-income worker ratio is 109; and the high-wage job to high-income worker ratio is 106. One thing to mention is that the overall average job-to-worker ratio does not equal to the average of wage-specific job-to-worker ratios, because the averages are calculated across block groups and when the jobs and worker are not evenly distributed across the block groups, the overall job-to-worker ratio may be smaller than the wage-specific job-to-worker. This explains why the low-wage, mid-wage, and high-wage job-to-worker ratios are 111, 109, and 106, while the overall average is only 103.

The job-to-worker ratios for all types of jobs are shown in Table 6.10. Overall, the change in job-to-worker ratios in the 12 scenarios follow a very similar pattern of change as accessibility. All types of job-to-worker ratios increase substantially in the 12 scenarios compared to the base scenario and the growth in the first quantile job-to-worker ratios is also the most substantial compared to other quantiles. This is consistent with our intuition that increasing job accessibility will result in similar percent increase in job-to-worker ratios when the employment and residential location patterns are fixed in the region. The similar pattern of change in accessibility and change in job-to-worker ratios are also related to the pretty evenly distributed jobs and workers in the region.

Table 6.10. Change of Job-to-Worker Ratio in the Twelve Scenarios

Overall Job-to-Worker Ratio Change								
	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	0	n.a.	0	n.a.	0	n.a.	0	n.a.
1st Quantile	90	1536%	88	1501%	96	1651%	95	1630%
Median	221	501%	219	496%	242	557%	244	562%
3rd Quantile	408	270%	411	274%	442	301%	449	308%
Max	115,984	446%	110,492	420%	134,763	534%	127,263	499%
Mean	375	264%	371	260%	412	300%	411	299%
	Scenario 5		Scenario 6		Scenario 7		Scenario 8	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	0	n.a.	0	n.a.	0	n.a.	0	n.a.
1st Quantile	62	1018%	61	1009%	66	1107%	66	1092%
Median	160	334%	159	332%	175	374%	174	373%
3rd Quantile	312	183%	313	184%	343	212%	342.	211%
Max	76,450	260%	72,906	243%	84,386	297%	82,123	287%
Mean	287	179%	287	179%	313	205%	315	206%
	Scenario 9		Scenario 10		Scenario 11		Scenario 12	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Min	0	n.a.	0	n.a.	0	n.a.	0	n.a.
1st Quantile	35	540%	35	534%	37	566%	36	561%
Median	103	179%	102	176%	109	197%	110	199%
3rd Quantile	222	102%	227	106%	242	120%	246	123%
Max	38,816	83%	35,014	65%	41,711	96%	41,060	93%
Mean	201	95%	205	99%	218	112%	225	119%

As the statistics show, there is not a big difference in wage-specific job-to-worker ratios in the study area, so the accessibility increase brought by ODRS results in similar percent increase in job-to-worker ratios. It can be imagined that if there is some significant mismatch between certain type of jobs and workers in a region, the impact of ODRS on job-to-worker ratio may not follow such an equitable pattern.

6.9 Conclusions

It is only several years since ride-sourcing emerged and attracted great attention, and the intrinsic nature of ODRS assembles some of the important features of automated vehicles that will become a core component of future sustainable transportation. Considering the growing impact of ODRS and its potential to be integrated with public transportation, understanding its accessibility and equity influence has important implications not only for now but also for the transition stage to an era of automated vehicles. This part of the dissertation quantifies fine-level impact of ODRS on job accessibility and transit service equity in the Puget Sound region.

The analytical results show that integrating ODRS with public transportation has substantial accessibility benefits. In the Puget Sound region, using ODRS for up to one mile around transit stops and using it to serve trips within two miles can improve the average block group level accessibility by about 100% (from 52 thousand to about 100 thousand). Using ODRS for up to two miles around transit stops or using it to serve single modal trips within four miles can increase the current average accessibility by more than 200%. Moreover, the potential accessibility increase for the areas with lowest current accessibility is mostly significant. The first quantile of block groups see an average accessibility increase by more than 400% in Scenarios 9 – 12, by more than 900% in Scenarios 5 – 8, and by more than 1400% in Scenarios 1 – 4. In Scenarios 9 – 12 that assume the most modest use of ODRS, the first quantile of block-group-level accessibility equals to the median in the base scenario. Another important finding from the analysis is that the potential accessibility increase is very evenly distributed across jobs/workers of different wage/income categories.

The results suggest that using ODRS to serve short trips either connecting transit stops or for single modal trips can substantially elevate the existing level of job accessibility. With the average wait time of ODRS ranging from 1 to 12 minutes, the results of different scenarios suggest that the availability of ODRS and the distance that ODRS can be used for are most important for affecting the extent to which accessibility can be improved, while the influence of wait time of ODRS is not so substantial. The accessibility improvement is most significant for areas with lowest existing accessibility and has no obvious difference across jobs/workers with different income levels.

An important limitation of the accessibility analysis in this dissertation is the lack of cost as a consideration. The impact of ODRS on job accessibility is estimated for block groups, which is fundamentally estimating the change to locational accessibility, so whether ODRS can really be used by an individual traveler is not considered in the estimation. Therefore, though the results suggest that the potential increase to job accessibility because of ODRS follows a very equitable pattern, different individuals may have different challenges to really use ODRS. For example, according to the 2014 Puget Sound regional travel survey data, only 47.4% of the persons from low-income households (annual income less than $25,000) have smartphones and another 10.0% are planning to get smartphones in 2014 (see Table 6.11). The smartphone ownership of low-income people is significantly lower than persons from mid-income and high-income households, which will definitely limit their ability to use ODRS. Moreover, the current cost of using ODRS like Uber and Lyft is much higher than using public transportation and ODRS is still mostly unaffordable for low-income people. Based on a simple estimation of the cost of using UberX according to the official Uber's webpage called "Uber Pricing by City" shows that using UberX for a trip of 1 mile, 2 miles, 3 miles, and 4 miles would cost $5.5, $7.4, $9.5, and $11.6 respectively in Seattle, without considering any price surging effects. This suggest that ODRS may not be affordable even for very short trips for low-income transit-dependent travelers. Therefore, estimating individual-level accessibility change because of ODRS and taking into considerations like cost and smartphone ownership is an important field for future research.

Table 6.11. Smartphone Ownership by Income Levels

		Has smartphone					
		Yes		Not yet, but plan to get one in 2014		No (and don't currently plan to get one)	
		Count	Share	Count	Share	Count	Share
Household annual income levels	Under $25,000	491	47.4%	103	10.0%	441	42.6%
	$25,000-$49,999	1,022	60.0%	100	5.9%	582	34.2%
	$50,000-$74,999	1,152	67.1%	89	5.2%	475	27.7%
	$75,000-$99,999	1,175	75.4%	73	4.7%	310	19.9%
	$100,000 or more	3,024	83.8%	94	2.6%	492	13.6%

Data Source: 2014 Puget Sound Regional Travel Survey Data
Note: 1,740 out of the 12,198 persons included in the survey data have missing values in their smartphone information, so are omitted in this statistical summary.

The analysis has several other limitations. First, the development of the scenarios only considers the change of travel distance and wait time of ODRS and does not consider future possible social, demographic and land use changes. Second, job accessibility is measured by a simple locational metric that has its limitation in reflecting real accessibility that an individual may experience. Also, the research only considers the accessibility benefits of integrating ODRS with public transportation and thus is neglecting the potential benefits of combining ODRS with walking, biking, or other travel options, which may underestimate the potential accessibility benefits of ODRS. Nevertheless, this work is the first attempt that quantifies the accessibility and equity benefits of ODRS and the developed methodology can be easily transferred to analyzing other regions or other considerations by modifying the assumptions.

CHAPTER 7. FINDINGS AND DISCUSSION

The dissertation examines the role of on-demand ride service (ODRS) in a multimodal transport system, explores factors related to the choice of ODRS and how to model travel mode choices of ODRS, and forecasts its potential impact on transport accessibility and equity. In this dissertation, ODRS refers to taxi and ride-sourcing services (e.g. Uber and Lyft) that provide point-to-point mobility service without requiring automobile ownership. ODRS can be considered as a transitional travel mode of automated vehicles, which are likely to become a core feature of future sustainable urban environments. With limited empirical data of ODRS available, the dissertation utilizes various public data sets in the United States to further the comprehensive understanding about ODRS. The dissertation generates new knowledge not only about this rapidly growing travel mode, but also about the multimodal nature of our transport systems, and explores methodological and implementational possibilities of modeling, forecasting, and visualizing impacts of disruptive transportation technologies.

This chapter summarizes most important findings from the dissertation and discusses their implications. The chapter is organized according to the results of several specific research questions under investigation: (1) what is the role of ODRS in a multimodal transport context; (2) why do people choose ODRS and how can we model the choice of ODRS in a travel demand forecasting process; and (3) what is the potential impact of ODRS on transport accessibility and equity.

7.1 The Role of ODRS in the Multimodal Travel Context

The first research question attempts to further the understanding about the role of ODRS. First, the users of ODRS including both taxi and ride-sourcing are examined to extract the socio-demographic characteristic of ODRS riders. Then a classification analysis is applied to the taxi trip data in New York City to reveal to what degree, taxi is competing with public transportation, versus complementing it or serving the first/last mile of transit. Then several regression models are developed to unravel the characteristics of places that generate different types of ODRS trips and the characteristics of places that have higher vs. lower ODRS trip generation. There are several important findings from the analytical results of the first research question.

7.1.1 Captive vs. Choice Users of ODRS

The socio-demographic and economic characteristics of taxi riders in New York City and of ODRS users nationwide suggests that ODRS has a role in serving transport-disadvantaged population and different markets including, both choice users and captive users who may have very different travel needs and behaviors. It is found that approximately 54.7% of the taxi trips are serving disabled, low-income, elderly, retired or unemployed people in New York City. The nationwide data does not have direct disability information, but has an indicator showing whether medical device is used in the trip which reflects to some extent whether the traveler has physical disability. About 37.2% of the ODRS riders nationwide are low-income, elderly, retired, unemployed people or people using medical devices. The fact that a large proportion of ODRS riders are transport-disadvantaged population confirm the paratransit role that ODRS has and reflects the potential of leveraging ODRS to improve transport equity.

The socio-demographic characteristics of taxi riders in New York are a somewhat different from that of ODRS riders on average as ODRS riders, according to the national data, consists of more well-educated people, and less female and elderly people. This be a result of the difference between conventional taxi riders versus ride-sourcing users. Currently the ride-sourcing service is only accessible via smart phones, so it creates barriers for equal access to the service, especially for those who are physically or economic disadvantaged. In the face of the rapidly increasing ride-sourcing service and emerging travel mode such as automated vehicles, it is important to consider how to improve physically or economically challenged people's access to ODRS. Providing subsidies for ODRS trips made by transport-disadvantaged population and providing incentives for ODRS companies to make their service requests easier, flexible, and better integrated with public transportation, are necessary. Launching policy and regulations to require at least certain portions of ODRS service to be wheelchair accessible or ADA compliant is the next step to improve equal access to ODRS. NYC has announced its goal of making 50% of its taxi fleet wheelchair accessible though currently only about 1.8% of taxis are wheelchair accessible (Donohue, 2013; Fitzsimmons, 2015).

7.1.2 Multimodal Connection between ODRS and Transit

Regarding the relationship between ODRS and fixed-route transit, the classification analysis of research question 1 reveals the multifaceted nature of the relationship between ODRS and transit. First, about forty percent of taxi trips in NYC are competing with public transit of high level of service. Another fifty percent of the taxi trips are serving areas with low level of transit service. There is also a significant proportion, about seven percent, of the taxi trips are made likely to serve the first/last mile of transit. Taxi is serving very different travel demand in different circumstances. Though this type of trip categorization cannot be validated with empirical data, the very distinct characteristics of the three types of taxi trips, to some extent, confirms the effectiveness of the categorization. The analysis substantiated the hypothesis that ODRS can be used to support the use of transit or fill in the gaps of the transit network, which further reveals the possibility of improving multimodal mobility and accessibility via integrating ODRS with transit. The implications from this analysis are also multi-faceted.

It is important to recognize the competing relationship between ODRS and public transportation and start to think about how transit should respond to the continuing growth of ODRS. The taxi trip categorization analysis uses the 2011 taxi trip data, but even back then, more than forty percent of taxi trips made are competing with good-quality transit service. Though the reason is not examined in the dissertation, but more convenient, more comfortable, and higher level of privacy are apparent advantages of ODRS compared to public transit. A recent study suggests that ride-hailing is taking away transit usage and is exacerbating roadway traffic congestion (Clewlow & Mishra, 2017), which would leave the travelers who cannot afford ride-hailing service with worse transit. It is true, to some extent, the growth of ODRS threatens the use of transit, but it also prompts the transit system to improve and to evolve. Just as the analysis in the dissertation suggests, ODRS could contribute to expanding the coverage areas of transit, so policy and practice should focus more on augmenting the synergistic effect and improving the multimodal connection between ODRS and transit. Rather than looking at private ODRS companies as competitors, seeking collaboration with these companies to improve demand-responsive transit, information provision, payment easement, and trip/route planning etc. is probably more effective for public transit operators to evolve in the long term.

The efficiency of a transit network not only depends on the transit system itself but also on how people can access the system. It is commonly accepted that walk sheds of transit are about 0.5 miles from a station, meaning people are willing to walk 0.5 miles to access transit. The multimodal connection between bike and transit has received more attention recently, mainly because biking can enlarge the catchment area of transit with low cost. However, biking has its limitation as an access mode to transit: biking is sensitive to weather, biking routes, and the number of people traveling, it has a distance limitation and it often requires the bus and train to have bike racks which have capacity constraints. ODRS, in contrast, does not have similar constraints.

The major constraint of taking ODRS as an access mode to transit should be its high cost compared to walking or biking. Based on the findings of research question 1, the clear majority of transit-extending taxi trips in New York City are shorter than two miles with the cost of less than ten dollars, indicating the distance/cost for which people are willing to travel by taxi to access transit under the current pricing mechanism. Though it is hard to predict the elasticity of taking a taxi to access transit, incentives to reduce the cost of ODRS trips connecting to transit should be considered as a strategy to improve the multimodal connectivity between ODRS and transit. Moreover, transit-extending taxi riders are found to use cash for payment much most frequently, compared to other taxi riders. King and Saldarriaga (2016) suggested that paying cash for taxi rides are associated with unbanked low-income populations. Combining payment methods of transit and ODRS might be effective to promote the integration of ODRS and transit. Given the substantial proportion of taxi trips taken to connect transit and the great potential of the taxi or other ODRS service such as automated vehicles, it is worthwhile to consider feasible and specific strategies and services to improve multimodal connectivity that promote transit use and elevate the convenience of multimodal travel.

Taxis share many similarities with ride-sourcing and automated vehicles, as they serve point-to-point, flexible route, and are based on request. Ride-sourcing is found to have replaced taxi trips mostly, but it also replaced some driving and transit trips (Rayle et al., 2016). Ridesharing and real-time information provision might be a key difference that distinguishes ride-sourcing or shared automated vehicles from traditional taxis. Encouraging more shared trips by ODRS is one approach to reduce some of the travel cost of ODRS that can facilitate more extensive use of ODRS, especially for transit-extending trips. Additionally, ridesharing largely depends on trip matching and trips starting or ending with transit stations can often be easily matched with other trips so strategies that encourage ridesharing connecting to transit can be one approach to encourage environmental friendly multimodal travel.

Although some built environment variables in the analysis, like employment density, median housing values, poverty rates, and land use, are correlated with some ‘pre-determined' characteristics of transit stations' locations, the results of the models reveal significant differences in the areas that the three types of taxi trips are serving. Generally, transit-competing taxi trips are serving areas with great subway/train availability, which are often highly developed downtown areas with high land values. Transit-extending and transit-complementing trips are serving areas on the opposite side, which are more likely to be peripheral areas with low density and lower land values. However, this finding indicates the complementary role that taxi has in serving areas where transit does not operate efficiently. The positive association between bus density and transit-extending trips indicates that in those areas taxis are probably replacing some bus trips. The different market segmentations that taxis serve are also related to the notion of ‘captive' versus ‘choice' users. Captive users are often defined as ‘transit dependent' travelers who may not have many travel options, while choice users are referred to ‘discretionary users' who can choose from several travel options (Giuliano, 2005; Jacques, Manaugh, & El-Geneidy, 2013; Polzin, Chu, & Rey, 2000). It is hard to distinguish these two types of users in transit-complementing and transit-extending trips, meaning that it is hard to know whether a taxi rider chooses to take a taxi because of a lack of other options or simply because the taxi service maximizes the traveler's utility. However, compared to transit-competing trips, transit-complementing and transit-extending are more likely to be made by captive users, and transit accessibility analysis would be necessary to further identify the areas where people must take the taxi because of lacking other options.

In summary, ODRS plays a critical paratransit role in providing mobility to transit-dependent people, especially those physically and economically disadvantaged. They can also support fixed-route mass transit and serving areas with insufficient transit service in some circumstances. The specific findings from this analysis reveal the direction of future policy and strategies regarding improving the multimodal connection between transit and ODRS, which has the potential to encourage transit use, improve mobility level across the board, and reduce the environmental impact of the transport system.

7.1.3 ODRS and Equity: Gap between Demand and Supply

The regression analysis in research question 1 suggests that there may exist a severe mismatch between ODRS supply and the potential demand for ODRS. The supply of ODRS is catering to wealthier places that have higher development density, which leads to more potential for it to compete with public transportation. However, there is substantial potential need for ODRS from transport-disadvantaged populations who live in areas with less ODRS supply. This is a result of the current unbalanced supply of ODRS directed solely by the private sector and implies the potential benefits of planning intervention to achieve more equity.

As the analysis suggests, ODRS plays an important role in serving transit-dependent travelers by providing more travel options and flexibility. However, the supply of ODRS is highly concentrated in high-density and wealthy places, such as Manhattan in New York City. The spatial distributions of taxi generation and ride-sourcing trip generation are very similar, except that places with higher concentration of well-educated people have higher ride-sourcing trip generation. The significant variation in the average daily pick-ups by ODRS at the block level shows how much these services vary spatially. Moreover, the places with high ODRS activities exhibit very similar socio-demographic characteristics. They are places with significantly higher median household incomes, much whiter population, and extremely high population density.

Although the causes for this phenomenon needs further examination, the spatially unequally distributed supply of ODRS is an obvious reason. ODRS companies always seek to maximize potential profit by providing more frequent and better services to places with wealthier residents and high densities of potential users. The use of ODRS must depend on their supply that completely determines the availability and level of service of ODRS. At places with lower supply level of taxi or TNC service, travelers will need to wait for longer for the vehicle to arrive, if the vehicle ever arrives. Currently, there is little incentive for those ODRS companies to provide service to areas where they see lower possibility of usage. Although ODRS does not require vehicle ownership, the current pricing mechanism makes it more expensive compared to public transportation, especially for relatively longer trips. This is an important reason for the fact that most transit-extending taxi trips are shorter than 2 miles found by Wang & Ross, (2017) . The relatively higher costs of ODRS is a barrier for more widely using it as a complementary mode to fixed-route transit and can also rule out the usage of certain population completely, a large proportion of whom might be transit-dependent.

The intense activities of ODRS in NYC have revealed the great potential of leveraging it to improve mobility, especially for transit-dependent population. Since fixed-route mass transit cannot cover every corner of the region, ODRS can serve the first and last mile of transit and is flexible enough to serve areas with lower density that do not support mass transit. However, the market-driven supply of ODRS will never follow an equitable pattern without appropriate and effective planning and policy intervention. The current supply of ODRS concentrate in high-density area with great land use mix, which are also areas that transit infrastructure resides, so it creates more possibility of conflicts between ODRS and transit. Planners need to play a more active role in addressing the unequal supply of ODRS by directing more service to areas that collect transport-disadvantaged people or areas with worse transit service. Just like the widely discussed problem of mismatch between transit supply and demand, ODRS, as a publicly available travel mode run by private companies, is confronting more serious a problem of demand-supply mismatch that brings more equity issues.

Some cities have already begun to implement incentive programs to encourage the use of ODRS to supplement their existing mass transit systems. Since last year, several Florida cities (Altamonte Springs, Lake Mary, Longwood, Maitland and Sanford) are subsidizing all Uber trip fares for up to 20% and subsidizing trips to/from SunRail stations for up to 25% (Dovey, 2017). Lyft and Amtrak have also formed a partnership to serve the first/last mile of travel to Amtrak stations (Amtrak Media Center, 2017). These new public-private partnership (P3) models are great examples showing the potential of integrating shared mobility with transit.

Cost and spatial distribution are the two main areas where cities could intervene to encourage more usage of shared mobility services. The Uber and Lyft P3 models are straight forward strategies: to incentivize private ODRS providers to improve the level of service in target areas and lower the cost by subsidizing some trips. There are several challenges in forming such P3 models, which may also vary significantly given different cities' scale, potential usage of shared mobility, and existing transit infrastructure.

First, the supply of ODRS is not fixed and is hard to be guaranteed, which make it very uncertain about what types of agreement can be reached between the public and private sectors. For example, the wait time of every ride-sourcing trip is different at different time of day and day of the year and since the ride-sourcing drivers are not hired as full-time workers, the supply is always fluctuating and cannot be easily expanded.

Therefore, the supply of ODRS cannot be set as fixed as transit and is also hard to be monitored, because it is not like fixed-route transit that have planned supply measured by headway and time span. This makes it hard for the public sector to step in, because to what degree the ODRS trips or infrastructure should be subsidized is often obscure. This might be the reason why the Florida cities agreed that they will subsidize all Uber trips but will subsidize more for trips starting or ending at transit stations, because the supply is always fluctuating.

Second, the access to using ODRS options is another challenge. Currently, ride-sourcing can only be accessed via smart phones, making it only available to certain sub-populations and may also have embedded equity issues. Improving the public acceptance of ODRS is challenging in the current setting of Uber and Lyft service because they require the travelers use credit cards for payment. It is thus important for public sector to comprehensively review available ODRS services and provide incentives and subsidies that can make most bang for the bucks to improve transit use and to serve transit-dependent population.

Third, it is important to make distinctions between incentivizing ODRS trips that complement transit versus the ones that replace transit, but how to identify the threshold needs to be further researched. The first research question of the dissertation found that although a significant proportion of taxi trips serve as a complementary mode to fixed-route transit, about half the trips directly compete with public transit trips. As other ODRS services grow, it is important for planners to adjust practice to promote the synergistic relationship between ODRS and transit rather than encourage ODRS to take away transit use. It is of course a challenging issue, especially because the supply of ODRS is fluctuating, as discussed. To what degree ODRS should be subsidized and incentivized and how to monitor the performance of such programs are important questions. Since the phenomenon is so new and little empirical data is available to develop evidence-based policy implications, incremental practice changes should be carried out with frequent evaluation and monitoring to accumulate relevant knowledge.

7.2 Mode Choice Modeling of ODRS

The second research question of the dissertation models the choice of on-demand ride service (ODRS) in a travel demand forecasting context. Modeling travel mode choices considering the availability of ODRS might be the first step to incorporate ODRS into normal transportation planning processes. This research question identifies the most important factors associated with the choice of ODRS and explores the appropriate methodology of modeling the choice of ODRS. Three models, including a multinomial logit (MNL) model, an extreme gradient boosting (XGB) model, and a random forest (RF) model, are applied to four household travel survey datasets, including the 2017 National Household Travel Survey (NHTS), and regional survey data from the New York region, the Puget Sound region, and the Delaware Valley region. The analytical results of this research question facilitate better understanding about what factors are associated with the choice of ODRS and whether machine learning can be used to improve travel mode choice modeling. The most important findings are summarized, and implications are discussed below.

7.2.1 Comparing Machine Learning with Statistical Modelling

People's travel mode choices intertwine with many different factors and may have noticeable changes as technological advances, new travel modes and new data sources are available. Modeling travel mode choice is a critical step in travel demand forecasting and may also face challenges and opportunities as new travel modes and data sources become available. The dissertation is among the limited number of studies that explore machine learning's application to travel mode choice modeling and has included a relatively comprehensive list of independent variables that are ready for practical use.

To compare the performance of different statistical and machine learning models, both overall and travel mode-specific training and testing errors of each model are examined. Also, to understand whether the models' performances are robust to data changes, each of the models are run 100 times to estimate the average training and testing errors. Table 7.1 provides a quick summary of the average testing errors of the models developed using the four different datasets (see Section 5.3 for more details). As can be seen from the table, the XGB and the RF surpass the MNL significantly regarding total errors for all four datasets. The MNL model can achieve an overall prediction accuracy of about 42%, 21%, 77%, and 56% for the four datasets; XGB can achieve 57%, 90%, 82%, and 68% respectively; while the RF model can always achieve the highest prediction accuracy for all four datasets and its prediction accuracy rates are 77%, 91%, 86%, and 78%. On average, the RF model is the best at predicting the choice of modes with smallest shares. Both XGB model and the RF model have stronger predictive power compared to the MNL model.

Table 7.1. Summary of Average Testing Errors by Models and by Regions

		Total	Car	Bike	Walk	Transit	ODRS
2017 NHTS	MNL	58%	59%	62%	41%	59%	72%
	XGB	43%	36%	55%	23%	40%	68%
	RF	23%	19%	26%	12%	25%	39%
New York	MNL	19%	10%	53%	28%	15%	29%
	XGB	10%	3%	30%	18%	6%	28%
	RF	9%	4%	41%	10%	6%	27%
Puget Sound	MNL	23%	13%	40%	47%	30%	83%
	XGB	18%	8%	32%	34%	29%	86%
	RF	14%	5%	50%	23%	35%	94%
Delaware Valley	MNL	44%	24%	78%	61%	48%	86%
	XGB	32%	11%	79%	36%	39%	87%
	RF	28%	11%	63%	31%	36%	85%

In terms of travel-mode specific errors, the two machine learning models also show better performance compared to the MNL model in most cases. Looking at the average errors of predicting the choice of ODRS, the XGB model surpasses the MNL model in the national data significantly and achieves similar level of error in the other three datasets. The RF model has significantly lower ODRS-specific prediction compared to the MNL model in the national data, has similar errors in the New York and Puget Sound data while has higher error rate in the Delaware Valley data. Except ODRS and biking whose small shares are smallest in all four datasets, both machine learning models have significantly lower prediction rates in predicting the choices of other modes in four datasets compared to the MNL model. This may suggest that thought machine learning models can generally achieve better prediction performance, they are more dependent on data volume and the small number of observations can have more influence on the performance of machine learning than the statistical model.

Regarding the explanatory power, the two machine learning models are not comparable to the MNL model, as their result cannot be quantified or directly interpreted. The XGB and the RF model are tree-based, so allow measuring independent variables' importance in determining mode choices, which are better than some other black box machine learning models. The MNL model directly contributes to understanding the relationships between people's travel mode choices and other factors and is thus very useful for variable selection and deriving policy implications.

Regarding the effort of developing and implementing the models, the machine learning models and the MNL model have different strengths and challenges. The advantage of machine learning is it has very little limitation on the data structure and model specification. For example, travel distance is included in the machine learning models and the variables' importance metrics have shown that travel distance is an important independent variable in predicting mode choices, but it cannot be included in the MNL model due to the correlation between travel time and distance. Also, though the machine learning model requires effort of tuning some hyperparameters to optimize the model's performance, the whole model fitting process requires less attention and effort compared to the MNL model that requires very careful model specification and testing to examine whether the statistical assumptions hold.

In contrast, the MNL model can easily avoid the overfitting issue. Especially for a very unbalanced dataset, the XGB model may not suffer from overfitting issue for all choices combined but may have overfitting issue when predicting the choice with small shares. For example, in this analysis, the hyperparameters of the XGB models are tuned by minimizing the multi-class predicting error and the overfitting issue is controlled at the whole dataset level. However, for the travel mode that only account for a very small mode share in the dataset, the tuned hyperparameters will likely result in overfitting for predicting the choices of that specific travel mode. The RF model has less concern in the overfitting issue and it also performs better in predicting the choice with small shares. Therefore, when developing machine learning models, attention needs to be paid to such issues and techniques such as resampling are always worth trying to make the dataset more balanced. Techniques that can handle choice-specific overfitting control need to be explored more.

Future mode choice modeling efforts should consider using machine learning techniques or integrating some machine learning techniques to conventional statistical modeling. The modeling results in the dissertation clearly demonstrates the advantages of machine learning models and the MNL model in different respects. Machine learning models have higher prediction accuracy than the MNL model. The MNL model allows intuitive interpretation that machine learning cannot surpass. An easy way to combine the advantages of the two types of models is to use MNL model as a before-hand variable selection and interpretation tool, while use machine learning to perform or improve the forecasting accuracy. Unbalanced data is a notorious problem in machine learning and the analytical results of dissertation suggests that when the smallest share of a dataset is larger than 5%, both the MNL model and the machine learning models can achieve significantly better performance. Of course, this may not hold true for other datasets, but avoiding extremely unbalanced datasets by collecting more samples for small share modes may be useful for improving models' performance in predicting travel mode choices. Machine learning models that perform well with unbalanced datasets are useful in this case. It is also found that the small number of observations may impact the performance of machine learning models more than the statistical model. Applying machine learning when the sample size is big enough may be able to achieve significantly better prediction accuracy than the statistical model.

Another challenge of applying machine learning to transportation modeling is how to effectively present the results of machine learning models to the public, communities, and decision-makers. The author thinks more training about the basics of machine learning and practical exploration are worthwhile to start to consider a wider application of this useful tool in the transportation field. Relevant training and workshop sessions and to include relevant courses in planning and transportation related curriculum may be good starting point.

Applying machine learning to the urban and transportation field may have broad implications and potentials, in addition to the analysis shown in this dissertation. First, there have been increasingly more studies that use machine learning on urban and transportation inquiries and the results have consistently suggest the strong prediction power of machine learning. Though to what extent machine learning can improve the performance of statistical models is not conclusive and definitely varies in different circumstances, machine learning has the potential to elevate forecasting accuracy generally, which can be applied to different subfields of urban studies, such as transportation planning, housing and land use policy, spatial analysis, economic development, etc. Second, machine learning is much more flexible in terms of the data type and format it can address and relationships between variables it can work with, compared to statistical models that often have strict priori assumptions and data format requirements. Now in the big data era, machine learning has the potential to generate new knowledge by revealing new patterns and new relationships among different data and information. The nature of being data-driven makes machine learning more powerful in the data abundant environment and is likely to generate knowledge in a more timely and flexible manner. Last but not least, machine learning plays a key role in shaping many new technologies and thus understanding machine learning is becoming an increasingly important component when related plans and policies need to be generated. For example, technologies such as automated vehicles and smart cities rely on machine learning algorithms significantly, understanding what the strength and deficiencies are in those technologies will become a critical step of analyzing and making related plans and policies. In sum, machine learning is shaping and influencing many disciplines and new technologies, and its influence on urban and transportation planning is starting to emerge. How to leverage machine learning for better forecasting, knowledge generating, and plan making is an important yet new topic in the field and needs to be researched more.

7.2.2 Factors Associated with the Choice of ODRS

The identification of factors associated with the choice of ODRS comes only from the MNL model as the machine learning models do not allow interpretation of the relationship between independent and dependent variables. For the four datasets that are modeled, including the 2017 NHTS data, and the regional travel survey data from the New York metropolitan area, the Puget Sound region, and the Delaware Valley region, similar sets of independent variables, including trip characteristics, personal/household treats, and neighborhood factors are found to be statistically significantly associated with people's mode choices. However, the relationships between some trip characteristics and the choice of ODRS are somewhat different in different models, while most personal/household and neighborhood variables are found to have consistent relationship with the choice of ODRS in the three regions.

Table 7.2 summarizes the relationship between the independent variables and the choice of ODRS in the four models to present a quick comparison. Though a set of independent variables are found to have consistent relationship with the choice of ODRS across the models, there is obvious inconsistency in the models' results. This is probably due to the geographical difference and might be a result of the small number of observations in the Puget Sound and Delaware Valley regions' data. In Table 7.2, all the variables that are found to be significantly associated the choice of ODRS in at least two models are marked, as they are more likely to be the ones that really and more universally influence people's choice of ODRS.

Travel time and travel cost are the two variables that are found to be negatively associated with the choice of all travel modes in all models, which is self-explanatory. Several trip factors are found to be associated with the choice of ODRS. Trips made for changing travel mode is found to be positively associated with the choice of ODRS in both the New York and the Puget Sound regions. This indicates that ODRS serves more multimodal travel demand and may fit into trips' gaps. ODRS trips are less likely to be made during peak hours and more likely to made late at night, which might be caused by traffic congestion during peak hours, peak-hour price surging, the unavailability of transit at night, safety considerations of biking or walking at night, or might be associated with special trip purposes such as recreation trips that happen more often at night.

Table 7.2. Summary of Factors Influencing the Choice of ODRS

	2017 NHTS		New York		Puget Sound		Delaware Valley	
	Trip Variables							
	Sign	Sig.	Sign	Sig.	Sign	Sig.	Sign	Sig.
Travel time	-	***	-	***	-	***	-	***
Trip cost	n.i.		Insig		-	***	-	*
Trip purpose: home	n.i.		Insig		Insig		n.i.	
Trip purpose: work	+	*	+	.	Insig		n.i.	
Trip purpose: recreation	+	***	Insig		Insig		n.i.	
Trip purpose: maintenance	n.i.		Insig		Insig		n.i.	
Trip purpose: change mode	n.i.		+	***	+	**	n.i.	
Loop trip	-	***	n.i.		n.i.		n.i.	
Home-based other trip	n.i.		n.i.		n.i.		-	**
Morning peak	-	***	-	***	Insig		Insig	
Evening peak	-	***	Insig		Insig		Insig	
Late night	+	***	Insig		+	***	+	**
Activity duration	n.i.		+	***	Insig		Insig	
Number of travelers	+	***	-	.	Insig		Insig	
	Personal / Household Variables							
Low income	+	***	+	***	Insig		-	.
High income	n.i.		+	***	Insig		Insig	
Disability	+	***	+	***	n.i.		+	***
Female	n.i.		+	**	Insig		Insig	
Younger than 18	n.i.		Insig		Insig		-	**
Elder than 65	-	***	Insig		Insig		+	**
Household size	-	**	-	***	-	*	-	***
Student	n.i.		-	*	Insig		n.i.	
Vehicles per capita	-	***	-	***	Insig		-	***
Driver's license	n.i.		n.i.		Insig		-	***
Use smart phone everyday	+	***	n.i.		n.i.		n.i.	
	Neighborhood Variables (Either origin or destination)							
Population density	+	***	+	***	+	***	+	***
Employment density	n.i.		+	***	+	*	+	***
Bus stop density	n.i.		+	***	n.i.		n.i.	
Subway station density	n.i.		+	***	n.i.		n.i.	
road density	n.i.		+	***	n.i.		n.i.	
Employment diversity	n.i.		+	***	n.i.		n.i.	
Employment-population balance	n.i.		Insig		+	**	n.i.	

Note: "+" indicates positive relationship; "- "indicates negative relationship; "Insig" = "Insignificant", means the variable is not statistically significant in the model result; and "n.i." = "Not Included", indicates that the variable is not included in the final model either due to data unavailability or model specification consideration. The bold font indicates that the variable is found to significantly associated with ODRS in at least two models out of four.

Regarding the personal and household characteristics, household size and vehicle ownership are the two variables that are consistently found to be negatively associated with the choice of ODRS in the three regions. This reflects the users of ODRS consist of more people in their earlier life stages or people who do not own vehicles. Disability is also found to be positively associated with the choice of ODRS in the national data and both the New York and the Delaware Valley regions. Though the New York City has the 'Access-A-Ride' program that reduces the cost of taxi trips for physically challenged population, this finding reflects a more universal dependency of disabled people on ODRS for mobility.

Population and employment density are found to be consistently and positively associated with the choice of ODRS in all models. This pattern probably results from the fact that ODRS is often more accessible in high-density areas. On one hand, this directly shows that ODRS trips are more likely to be made in densely developed areas that often have higher density, better land use mix, and better transit services. On the other hand, this may imply that people's choice of ODRS can be substantially influenced by the supply of ODRS services that directly determines the wait time of ODRS, since currently ODRS services including taxis and ride-sourcing are more available in downtown than in suburban areas Associated independent variables such as job-housing balance, employment diversity, and transit facility density are reflecting the characteristics of such areas. This point needs more research to be warranted

7.2.3 *Incorporating ODRS into Travel Demand Forecasting: The Way Ahead*

Modeling the mode choice of ODRS is the first step of forecasting and estimating the potential impact of ODRS and incorporating it into a normal transportation planning process. The mode choice modeling analysis in this dissertation reveals that the factors influencing people's choice of ODRS are like those factors that are known to influence choice of other modes. Also, the strong prediction power of using machine learning for travel mode choice modeling reveals the potential of leveraging new analytics method to improve travel demand forecasting. However, to really incorporate ODRS into travel demand modeling still faces several challenges and calls for more research and practice exploration.

First, the existing public household travel survey data are designed and collected in a conventional way that may not be able to reveal factors that influence people's choice of new travel mode like ODRS. The first necessary effort to incorporate ODRS into transportation planning is to collect more data. The necessity of collecting more data of ODRS trips lies in twofold. First, sufficient number of observation of ODRS trips is the foundation to develop solid travel demand forecasting models that considers the availability of ODRS. As shown by the mode choice modeling analysis in the dissertation, the best accuracy of predicting ODRS mode choices for the New York region is about 73.3% and sharply drops to only 17.2% for the Puget Sound region and 15.4% for the Delaware Valley region because of the small number of observations of trips made by ODRS in the two regions. ODRS only accounts for 0.3% of the trips in the national data, 0.8% in the New York region, and only 0.2% in the Puget Sound and Delaware Valley regions. Even the RF model that is known for good at dealing with unbalanced data has limited power in predicting such rare cases. Even though the sampling techniques that can make the performance of the models improve significantly, the insufficient observations of trips made by ODRS impairs our ability to thoroughly understand the underlying causal relationships. When the number of observations is too small, the estimation of MNL models may become biased and fail to reveal the true relationship between dependent and independent variables, making it very hard to derive correct and effective policy implications.

Another reason that makes the collection of ODRS related data critical is related to the uniqueness of this travel mode. Though the travel mode choice modeling analysis in the dissertation shows that a similar set of independent variables is associated with the choice of ODRS and other conventional travel modes, it could simply be a result of the limitation that only those conventionally known variables are available from the datasets. Considering the uniqueness of ODRS, it is very likely that some distinct factors may have important influence on the choice of ODRS but are yet included in any of the survey data that is currently available. For example, the use of ride-sourcing must rely on using credit cards, so whether a traveler has credit cards determines whether ride-sourcing is available. Such factor is not available in most of the existing travel survey data, so such obvious causal relationship cannot be captured in travel demand modeling. A successful incorporation of ODRS into travel demand forecasting thus calls for more data collection and better-designed data collection that include more new factors and questions that may not have not been considered in conventional travel survey data.

Another challenge that is important for modeling the travel behaviors of ODRS travelers is about how to better model multimodal travel. The analytical results of the dissertation show clearly that ODRS plays an important role in serving transport-disadvantaged population, varied market segmentations, and multimodal travel connecting public transportation or other travel modes. The rapid growth of ODRS users and the potential ODRS service provided by automated vehicles might be able to further strengthen such patterns. Therefore, how to model the potential shift to more multimodal travel that are likely to be more discretionary, flexible, and better integrated with active modes like walking and biking will become a challenge, especially for modeling the choice of ODRS. Collecting related empirical data and explore methodological feasibility of modeling the multimodal travel and trip chaining effects is important.

Finally, the choice of ODRS may be largely affected by the supply side that implies there might be great room in which policies and technologies can intervene to influence people's choice of ODRS. As the analysis shows, trips made by ODRS are positively associated with densely developed areas with mixed land use patterns and better transit services which are mostly likely downtown areas. This is probably associated with the fact that travelers can fit ODRS more easily into a multi-modal trip in those areas as more travel options are available but may also be closely related to the fact that ODRS is often more available in those areas. How to appropriately and effectively intervene to direct more ODRS to areas that have higher concentration of transport-disadvantaged population is an important question. To what extent the current choice patterns of ODRS will change when the supply or access to ODRS changes is a following question. More research, practice, and data collection effort need to be realized to gradually achieve a more comprehensive understanding about what is the best way to incorporate ODRS into travel demand forecasting.

7.3 Impact of ODRS on Transport Accessibility and Equity

This third part of the dissertation investigates the potential improvement of job accessibility because of ODRS, responding to the recent fast growth of ODRS users and the growing concern that ODRS is taking away transit usage. Quantifying and measuring the potential accessibility benefits can help reveal strategies to maximize the synergistic effect between ODRS and fixed-route transit. Several important findings are summarized and discussed as follows.

7.3.1 Realizing the Substantial Accessibility Benefits of ODRS

The analytical results of the third research question show that integrating ODRS with public transportation has substantial accessibility benefits. In the Puget Sound region, using ODRS for up to one mile around transit stops and using it to serve trips within two miles can improve the average block group level accessibility by about 100% (from 52 thousand to about 100 thousand). Using ODRS for up to two miles around transit stops or using it to serve single modal trips within four miles can increase the current average accessibility by more than 200%. Moreover, the potential accessibility elevation for the areas with lowest accessibility is mostly significant. The first quantile of job accessibility increases by more than 400% in Scenarios 9 – 12, by more than 900% in Scenarios 5 – 8, and by more than 1400% in Scenarios 1 – 4. In Scenarios 9 – 12 that assume the most modest use of ODRS, the first quantile of accessibility equals to the median in the base scenario.

Such considerable accessibility increase because of ODRS is associated with the huge gap between car and transit accessibility. On one hand, currently transit can only serve travelers within 0.5 mile from a transit station but using ODRS can enlarge the catchment area of transit which can significantly increase the number of travelers that transit can serve and make more jobs more accessible by using transit. On the other hand, in the scenarios, we assume that ODRS can be used to serve trips shorter than 2 miles, 3 miles, and 4 miles respectively, similar as paratransit, which is able to reduce the gap between car and transit accessibility for short distance trips. This means that assuming ODRS is available everywhere, any jobs that are located within 2, 3, or 4 miles will be considered as easily accessible, meaning that the catchment areas of employment are enlarged significantly.

The results reveal the huge accessibility benefits of leveraging ODRS to provide better access to transit and to provide point-to-point mobility service. The accessibility benefits that are quantified can be easily monetized and used to compare the cost-effectiveness of using ODRS to provide certain level of accessibility versus using other modes. Therefore, it provides a base for implementing strategies to integrate ODRS with transit by providing incentives or subsidies to ODRS for targeted areas. Moreover, the analysis also suggests a way to identify target areas of leveraging ODRS to improve accessibility. For areas that have low accessibility currently and higher concentration of transit-dependent populations, subsidizing ODRS trips to/from a transit station for short distances can be very cost effective to improve mobility and accessibility for captive transit users. Especially for areas with low-density and cannot support mass transit, ODRS provides new opportunity of enhancing transit accessibility and equity.

7.3.2 Leveraging ODRS to Improve Transport Equity

Another important finding from the analysis is that the potential accessibility increase is very evenly distributed across jobs/workers of different wage/income categories. Though the region has larger shares of higher income workers/jobs than mid- or low- income workers/jobs, the percent growths of accessibility are similar, and the percent increase of job-to-worker ratios are also similar across the three types. Changes in wait time do not seem to influence the accessibility benefits significantly, and the distance that ODRS can travel is the dominating factor determining to what extent accessibility can be improved. This indicates that if ODRS is available, the potential accessibility benefits will be almost equally distributed in the region.

Three main barriers of using ODRS may make the access to ODRS not equal. First, in the current market mechanism, the supply of ODRS is completely directed by private companies seeking profit maximization. ODRS is often not available everywhere in the region. For the areas without efficient ODRS, the existence of ODRS in other areas may be exacerbating accessibility inequality across different spatial locations. Second, the current cost of using ODRS like Uber and Lyft is much higher than using public transportation and ODRS is still mostly unaffordable for low-income people. Based on a simple estimation of the cost of using UberX according to the official Uber's webpage called "Uber Pricing by City" shows that using UberX for a trip of 1 mile, 2 miles, 3 miles, and 4 miles would cost \$5.5, \$7.4, \$9.5, and \$11.6 respectively in Seattle, without considering any price surging effects. This suggest that ODRS may not be affordable even for very short trips for low-income transit-dependent travelers. Third, there is also social and perceptual barriers for some population groups to use ODRS. Currently, ride-sourcing service is mostly only accessible on smart phones, travelers need to learn by themselves about how to use those ride-sourcing apps, and travelers are only allowed to pay with credit cards, and so forth. Moreover, low-income people generally have much smaller proportion of smartphone ownership, which make ODRS perceived as unavailable. Even when ODRS becomes available for most areas, these barriers will keep making the access to using ODRS unequal.

The contrast between the huge potential accessibility and equity benefits of ODRS and the obvious barriers of using ODRS equally reveals the role that practical and policy intervention has. Potential strategies of realizing the accessibility and equity benefits of ODRS include: (1) leverage accessibility analysis to identify target areas that have higher concentration of transit-dependent population and high potential accessibility benefits of integrating ODRS with public transportation; (2) incentivize ODRS companies to provide higher level of service around transit stations and/or in areas with higher concentration of transit-dependent population; (3) subsidize ODRS trips that connect to transit stops and/or subsidize short distance ODRS trips for target population groups and/or target areas; (4) enhance equal access to using ODRS across different population groups by reducing the social, economic, and perceptual barriers of using ODRS. Implementing these strategies must rely on sustaining collaboration between public agencies and private ODRS companies, which call for a more active role of transportation planners, engineers, government agencies, and transit operators to initiate such efforts.

The finding that ODRS can provide huge accessibility and equity benefits even when it is only used for short distance trips suggests the importance to initiate practical and policy efforts to realize such benefits. Implementing strategies such as enhancing equal access to ODRS and subsidizing ODRS trips for low-income transit-dependent travelers has to rely on a more active role of planners, government agencies, and transit operators to initiate building partnership with private ODRS providers. The current social, economic, perceptual barriers to using ODRS is a main challenge for providing equal access to ODRS and will continue to be the main challenge when ODRS is provided by automated vehicles, and thus should be researched more. An important field to extend this research is to conduct individual-level accessibility analysis taking into considerations like cost and smartphone ownership etc., to understand the barriers of using ODRS for transport-disadvantaged population groups.

CHAPTER 8. CONCLUSIONS

The emergence of innovative mobility services, such as bike-sharing, car-sharing, ride-sourcing, e-hailing, personalized public transit, and virtual mobility have provided travelers with unprecedentedly wide range of modality options for fulfilling their daily mobility needs and have brought new questions to sustainable transportation planning. As part of the phenomenon known as Mobility as a Service (MaaS), on-demand ride service (ODRS) has been acquiring increasingly larger market shares. ODRS is publicly available, provides point-to-point service, and does not require vehicle ownership, making it unique and different from both conventional private and public mobility options. By providing point-to-point mobility options and easily fitting into the gaps of transit, ODRS also has the potential to elevate mobility and accessibility generally and particularly for the transit-dependent travelers. In the US context, the vast difference between the mode shares of driving versus alternative transportation modes has not only resulted in the lack of capacity on most roads and highways, it has spawned the lack of funding for constructing infrastructure of all other modes. Urban accessibility and mobility are impaired, on one hand, by the increasing level of roadway traffic congestion, and on the other hand, by the lack of travel options especially for the physically or economically disadvantaged people. The emergence and fast growth of ODRS reveals new possibilities to significantly improve the transport benefits for larger population and make the transport system more inclusive.

This dissertation investigates several key questions surrounding ODRS, including its role in the multimodal transport context, how to model the choice of ODRS, and its potential impact on transport accessibility and equity. The dissertation focuses on examining taxi and ride-sourcing, because they are the two main forms of ODRS that provides everyday mobility. The dissertation uses assorted analytical approaches, including regression analysis, spatial analysis, discrete choice modeling, machine learning, scenario analysis, and so forth, to analyze multiple data sources of various formats, including household travel survey data, GPS trip data, general transit feed specification data, demographic and employment data, built environment data, etc. This dissertation attempts to further the comprehensive understanding about ODRS and its potential impact and explore the methodological possibilities of incorporating ODRS into normal transportation planning processes.

The dissertation reveals the important role that ODRS has in serving transport-disadvantaged populations, the varied market segmentations that ODRS is serving (captive vs. choice users), the current mismatch between ODRS demand and supply and related equity issue, the substantial potential accessibility and equity benefits of ODRS, the challenge of incorporating ODRS into travel demand modeling when data is inadequate, and the advantage of applying machine learning to travel mode choice modeling. These findings unveil many possibilities about how ODRS can be leveraged to elevate transport mobility, accessibility, and equity all over the board and reveal the potential room for improvement where planners, public sectors, and decision makers can play a role.

The implications can be classified into three general and intertwined groups, which are (1) reduce the mismatch between ODRS supply and demand by incentivizing ODRS companies to provide more equally distributed services, subsidizing certain types of ODRS trips, and reducing various barriers of using ODRS, to promote more equitable, convenient, and efficient use of ODRS; (2) improve overall transport mobility, accessibility, and equity by enhancing the multimodal connection between ODRS and other travel modes, especially transit and walking and biking, to encourage the shift to more active and sustainable travel; and (3) start to realize the collection and publication of all sorts of ODRS-related data and explore available computational and analytical methods to enhance solid research and modeling work of ODRS, which is critical for improving transportation planning in the era of shared mobility and automated vehicles.

A great amount of attention has been paid to ODRS since the recent rapid growth in ride-sourcing users. Nevertheless, a lot of discussions surrounding ODRS is about its "unfair" competition with traditional taxis and its replacement of public transportation. Ride-sourcing differs from traditional taxis in several major ways, such as it provides better real-time information, matches drivers and riders efficiently, lowers the cost of taxis to some extent, provides more ride-sharing opportunities, and has more flexible pricing mechanism that is not constraint by ordinance as the traditional taxi is. It is true that these new features of ride-sourcing and the improved service of ride-sourcing compared to traditional taxis make ride-sourcing more appealing and threatens the traditional taxi industry. Nevertheless, it is also true that people benefit from the improved service of ride-sourcing, which may be because of the improved convenience of the service, the lowered cost, and probably the easier payment method. Though this dissertation does not provide in-depth comparison between users of ride-sourcing vs. traditional taxis, the analysis suggests some fundamental similarities in users' characteristics and places that have high concentration of both types of services. The dissertation also suggests the important role that ODRS, including both ride-sourcing and taxis, has in improving mobility and accessibility especially for transport-disadvantaged population.

Therefore, even though the traditional taxi industry is experiencing some shocks brought by ride-sourcing, probably we have to admit that the rise of ride-sourcing has prompt regulators to rethink about the ODRS industry. Changes are happening to the traditional taxi industry as a response and the ODRS generally is evolving and is being improved. For some other places, taxi vehicles are equipped with GPS and also have mobile Apps that provide better real-time information (Estes, 2015; Poon, 2015). For some places, regulators are thinking about loosening the taxis' pricing ordinance (Farren, Koopman, & Mitchell, 2017; Wear, 2018). These new trends in the practice reflects improvement to the traditional taxi industry as a response to the recent shock due to ride-sourcing and though it may take, it is imaginable that services of traditional taxis and ride-sourcing may become more and more similar to each other or even converge eventually.

In contrast, there are also some necessity for regulators and decision makers to initiate effort on prompting the ride-sourcing industry to be improved to serve transport-disadvantaged populations more effectively. There has been some criticism about how ride-sourcing service fail to serve disabled people, mostly because of the lack of vehicles equipped to handle wheelchairs and motorized scooters (Kunkle, 2018; Sorrel, 2015). In addition to incentivizing more ride-sourcing drivers to equip their vehicles to better serve disabled people, it is also necessary for the public sector to collaborate with private ride-sourcing companies to direct more ODRS supply to areas with high concentration of transportation-disadvantaged population. Given the great potential benefits of ODRS and its anticipated growing market shares, removing barriers to using ODRS and making the service equal to different population groups will become a core factor in determining future transport equity.

Though automated vehicles will be able to provide higher level of mobility and accessibility options compared to the current form of ODRS. Ride-sourcing and taxis can provide drivers' assistance to transport-disadvantaged population, which is an important advantage compared to other travel options but has been mostly neglected so far. Transport-disadvantaged may have very specific physical or cognitive constraints that impact their use of transportation services. It is often beyond the funding and financing capacity of places, especially small urban and rural places, to provide extensive paratransit service to serve special population groups. Leveraging the flexibility of ODRS and its potential advantage in providing drivers' assistance to special groups of travelers, may be able to significantly improve the mobility and accessibility level of these travelers. More research is needed to identify the gaps in needs of these travelers and the availability of ODRS in different types of places, including urban, small urban, and rural areas. Also, more policy research is needed to identify pathways to encourage providing ODRS to vulnerable places and populations.

A gradual but substantial change is happening to our transport system as disruptive transportation technologies, such as ride-sourcing, shared mobility, and automated vehicles emerge. Within the next several decades or even sooner, we are likely to see very different patterns of travel behaviors, new options of travel modes, and shifting paradigm of urban transportation planning. How to facilitate smoother and more sustainable transition to new transportation technologies when the future is hard to predict? Proactive thoughts and solid research on emerging trends and newly revealed issues are important. This dissertation is among the early work that examines multiple questions surrounding ODRS, aiming to contribute to knowledge accumulation about this new travel mode even when empirical data are not abundant. The dissertation reveals the characteristics of ODRS, unveils its potential in promoting more inclusive and sustainable travel, and emphasizes the current issues and challenges of using and planning for ODRS. More importantly, the dissertation attempts to arouse more attention to this emerging travel mode that has the potential to change the way we travel and lead to more sustainable urban futures as the era of shared mobility and automated vehicles has already come.

REFERENCES

Aarhaug, J. (2014). *Taxis as urban transport*. Retrieved from http://trid.trb.org/view.aspx?id=1308462

Alsnih, R., & Hensher, D. A. (2003). The mobility and accessibility expectations of seniors in an aging population. *Transportation Research Part A: Policy and Practice*, *37*(10), 903–916.

Altshuler, A., Womack, J. P., & Pucher, J. R. (1980). The urban transportation system: Politics and policy innovation. *Journal of Economic Literature*, *18*(3).

Amtrak Media Center. (2017, August 1). Amtrak and Lyft Announce Rideshare Partnership. Retrieved December 6, 2017, from https://media.amtrak.com/2017/08/amtrak-lyft-announce-rideshare-partnership/

Austin, D., & Zegras, P. C. (2011). *The taxicab as public transportation in Boston*. Massachusetts Institute of Technology.

Bachand-Marleau, J., Larsen, J., & El-Geneidy, A. (2011). Much-Anticipated Marriage of Cycling and Transit. *Transportation Research Record: Journal of the Transportation Research Board*, *2247*, 109–117. https://doi.org/10.3141/2247-13

Bates, J., Polak, J., Jones, P., & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2), 191–229.

Ben-Akiva, M., & Lerman, S. R. (1979). Disaggregate travel and mobility choice models and measures of accessibility. *Behavioural Travel Modelling*, 654–679.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.

Benekohal, R. F., Michaels, R. M., Shim, E., & Resende, P. T. (1994). *Effects of Aging on Older Driver's Travel Characteristics*. Retrieved from https://trid.trb.org/view.aspx?id=413772

Best, H., & Lanzendorf, M. (2005). Division of labour and gender differences in metropolitan car use: An empirical study in Cologne, Germany. *Journal of Transport Geography*, *13*(2), 109–121. https://doi.org/10.1016/j.jtrangeo.2004.04.007

Bhat, C. R. (1997). Work travel mode choice and number of non-work commute stops. *Transportation Research Part B: Methodological*, *31*(1), 41–54. https://doi.org/10.1016/S0191-2615(96)00016-1

Bhat, C. R. (1998a). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, *32*(7), 495–507.

Bhat, C. R. (1998b). Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological*, *32*(6), 361–371.

Bhat, C. R. (2000). Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science*, *34*(2), 228–238.

Bhat, C. R. (2003). Random utility-based discrete choice models for travel demand analysis. *Transportation Systems Planning: Methods and Applications*, *10*, 1–30.

Biagioni, J. P., Szczurek, P. M., Nelson, P. C., & Mohammadian, A. (2008). Tour-based mode choice modeling: using an ensemble of (un-) conditional data-mining classifiers. Retrieved from http://www.biagioni.net/papers/biagioni-trb09.pdf

Boarnet, M. G., Giuliano, G., Hou, Y., & Shin, E. J. (2017). First/last mile transit access as an equity planning issue. *Transportation Research Part A: Policy and Practice*, *103*, 296–310.

Böcker, L., Dijst, M., & Prillwitz, J. (2013). Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. *Transport Reviews*, *33*(1), 71–91.

Bowman, J. L., & Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, *35*(1), 1–28.

Burkhardt, J. E. (1999). Mobility changes: Their nature, effects, and meaning for elders who reduce or cease driving. *Transportation Research Record: Journal of the Transportation Research Board*, (1671), 11–18.

Burkhardt, J. E., Berger, A. M., Creedon, M., & McGavock, A. T. (1998). Mobility and independence: Changes and challenges for older drivers. *Prepared by Ecosometrics, Inc., for the US Department of Health and Human Services and the National Highway Traffic Safety Administration*.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009a). Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews*, *29*(3), 359–395.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009b). The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part A: Policy and Practice*, *43*(5), 548–559. https://doi.org/10.1016/j.tra.2009.02.001

Carp, F. M. (1988). Significance of mobility for the well-being of the elderly. *Transportation in an Aging Society: Improving Mobility and Safety of Older Persons*, *2*, 1–20.

Celikoglu, H. B. (2006). Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. *Mathematical and Computer Modelling*, *44*(7), 640–658.

Cervero, R. (1994). Transit-based housing in California: evidence on ridership impacts. *Transport Policy*, *1*(3), 174–183. https://doi.org/10.1016/0967-070X(94)90013-2

Cervero, R. (1997). *Paratransit in America: Redefining mass transportation*. Greenwood Publishing Group.

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment*, *7*(4), 265–284. https://doi.org/10.1016/S1361-9209(01)00024-4

Cervero, R. (2007). Transit-oriented development's ridership bonus: a product of self-selection and public policies. *Environment and Planning A*, *39*(9), 2068–2085.

Cervero, R., & Duncan, M. (2003). Walking, bicycling, and urban landscapes: evidence from the San Francisco Bay Area. *American Journal of Public Health*, *93*(9), 1478–1483.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Cetin, T., & Deakin, E. (2017). Regulation of taxis and the rise of ridesharing. *Transport Policy*. https://doi.org/10.1016/j.tranpol.2017.09.002

Chang, H., Tai, Y., & Hsu, J. Y. (2009). Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining*, *5*(1), 3–18.

Chang, Y.-C. (2013). Factors affecting airport access mode choice for elderly air passengers. *Transportation Research Part E: Logistics and Transportation Review*, *57*, 105–112.

Chen, C., Gong, H., & Paaswell, R. (2008). Role of the built environment on mode choice decisions: additional evidence on the impact of density. *Transportation*, *35*(3), 285–299.

Chen, L., Mislove, A., & Wilson, C. (2015). Peeking Beneath the Hood of Uber. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference* (pp. 495–508). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2815681

Chorus, C. G., Molin, E. J., & Van Wee, B. (2006). Travel information as an instrument to change cardrivers' travel choices: a literature review. *EJTIR, 6 (4), 2006*. Retrieved from http://repository.tudelft.nl/view/ir/uuid:098105c4-612e-4814-93b9-bcd34aa71d24/

Christoforou, Z., Milioti, C., Perperidou, D., & Karlaftis, M. G. (2012). Investigation of taxi travel time characteristics. *Advances in Transportation Studies*, (27).

Clewlow, R. R., & Mishra, G. S. (2017). *Disruptive transportation: the adoption, utilization, and impacts of ride-hailing in the United States*. Research Report–UCD-ITS-RR-17.

Clifton, K., & Muhs, C. (2012). Capturing and Representing Multimodal Trips in Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, *2285*, 74–83. https://doi.org/10.3141/2285-09

Dalvi, M. Q., & Martin, K. M. (1976). The measurement of accessibility: some preliminary results. *Transportation*, *5*(1), 17–42.

Dias, F. F., Lavieri, P. S., Garikapati, V. M., Astroza, S., Pendyala, R. M., & Bhat, C. R. (2017). *A Behavioral Choice Model of the Use of Car-Sharing and Ride-Sourcing Services*.

Dissanayake, D., & Morikawa, T. (2002). Household travel behavior in developing countries: nested logit model of vehicle ownership, mode choice, and trip chaining. *Transportation Research Record: Journal of the Transportation Research Board*, (1805), 45–52.

Dissanayake, D., & Morikawa, T. (2010). Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference Nested Logit model: case study in Bangkok Metropolitan Region. *Journal of Transport Geography*, *18*(3), 402–410. https://doi.org/10.1016/j.jtrangeo.2009.07.003

Dongmei, L. I. U., Tongyan, Q. I., Zhang, K., & Yanmei, G. U. O. (2009). Beijing Residents' Travel Time Survey in Small Samples. *Journal of Transportation Systems Engineering and Information Technology*, *9*(2), 23–26.

Donohue, P. (2013). 50% of cabs wheelchair accessible by 2020. *NY Daily News*. Retrieved from http://www.nydailynews.com/new-york/50-cabs-wheelchair-accessible-2020-article-1.1539688

Dovey, R. (2017). 5 Florida Cities Team Up to Subsidize Uber Rides. Retrieved December 6, 2017, from https://nextcity.org/daily/entry/five-florida-cities-subsidize-uber-rides

Espín-Noboa, L., Lemmerich, F., Singer, P., & Strohmaier, M. (2016). Discovering and Characterizing Mobility Patterns in Urban Spaces: A Study of Manhattan Taxi Data. Retrieved from https://pdfs.semanticscholar.org/7bcd/2cbd7896f24b372944db70cdd5b6ef989dd9.pdf

Estes, A. C. (2015). California Taxi Industry Finally Has Its Own App, Only Four Years Too Late. Retrieved June 26, 2018, from https://gizmodo.com/california-taxi-industry-finally-has-its-own-app-only-1749441764

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*(1), 87–114. https://doi.org/10.3141/1780-10

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association*, *76*(3), 265–294.

Ewing, R., Schroeer, W., & Greene, W. (2004). School location and student travel analysis of factors affecting mode choice. *Transportation Research Record: Journal of the Transportation Research Board*, (1895), 55–63.

Fagnant, D. J., & Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, *40*, 1–13.

Fan, Y. (2012). The planners' war against spatial mismatch: Lessons learned and ways forward. *Journal of Planning Literature*, *27*(2), 153–169.

Fang He, & Shen, Z.-J. M. (2015). Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C: Emerging Technologies*, *58*, 93–106. https://doi.org/10.1016/j.trc.2015.06.023

Farber, S., Morang, M. Z., & Widener, M. J. (2014). Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, *53*, 149–159.

Farren, M., Koopman, C., & Mitchell, M. (2017, May 25). Ridesharing vs. Taxis: Rethinking Regulations to Allow for Innovation. Retrieved June 26, 2018, from https://www.mercatus.org/publications/ridesharing-vs-taxis-rethinking-regulations-allow-innovation

Fitzsimmons, E. G. (2015, August 31). Tomorrow Arrives for New York City's Yellow Cab Standard. *The New York Times*. Retrieved from https://www.nytimes.com/2015/09/01/nyregion/taxi-of-tomorrow-becomes-citys-yellow-cab-standard.html

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation*, *35*(1), 37–54.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Gärling, T. (1998). Behavioural assumptions overlooked in travel choice modelling. *Travel Behaviour Research: Updating the State of Play*, 3–18.

Gaver Jr, D. P. (1968). Headstart strategies for combating congestion. *Transportation Science*, *2*(2), 172–181.

Georggi, N. L., & Pendyala, R. M. (2000). *An analysis of long-distance travel behavior of the elderly and the low-income*. University of South Florida. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec026/02_georggi.pdf

Geurs, K. T., & Wee, B. van. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, *12*(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Gilbert, G., & Samuels, R. E. (1982). *The Taxicab: An Urban Transportation Survivor*.

Giuliano, G. (2005). Low income, public transit, and mobility. *Transportation Research Record: Journal of the Transportation Research Board*, (1927), 63–70.

Golob, T. F. (2000). A simultaneous model of household activity participation and trip chain generation. *Transportation Research Part B: Methodological*, *34*(5), 355–376. https://doi.org/10.1016/S0191-2615(99)00028-4

Golub, A., & Martens, K. (2014). Using principles of justice to assess the modal equity of regional transportation plans. *Journal of Transport Geography*, *41*, 10–20.

Grengs, J. (2010). Job accessibility and the modal mismatch in Detroit. *Journal of Transport Geography*, *18*(1), 42–54.

Gupta, S., Vovsha, P., & Donnelly, R. (2008). Air passenger preferences for choice of airport and ground access mode in the New York City metropolitan region. *Transportation Research Record: Journal of the Transportation Research Board*, (2042), 3–11.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, *25*(2), 73–76.

Hensher, D. A., & Ton, T. T. (2000). A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E: Logistics and Transportation Review*, *36*(3), 155–172.

Hess, S., Bierlaire, M., & Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, *39*(2), 221–236.

Hildebrand, E. D. (2003). Dimensions in elderly travel behaviour: a simplified activity-based model using lifestyle clusters. *Transportation*, *30*(3), 285–306.

Hochmair, H. H. (2015). Spatio-Temporal Pattern Analysis of Taxi Trips in New York City. Retrieved from http://docs.trb.org/prp/16-6511.pdf

How Much Does Uber Cost versus Taxis? (n.d.). Retrieved October 8, 2016, from https://techboomers.com/t/uber-cost

How Much More Can Ride-Sharing Services Grow in the US? - eMarketer. (2016). Retrieved October 7, 2016, from http://www.emarketer.com/Article/How-Much-More-Ride-Sharing-Services-Grow-US/1013963

Ipingbemi, O. (2010). Travel characteristics and mobility constraints of the elderly in Ibadan, Nigeria. *Journal of Transport Geography*, *18*(2), 285–291.

Jackson, W. B., & Jucker, J. V. (1982). An empirical study of travel time variability and travel choice behavior. *Transportation Science*, *16*(4), 460–475.

Jacques, C., Manaugh, K., & El-Geneidy, A. M. (2013). Rescuing the captive [mode] user: an alternative approach to transport market segmentation. *Transportation*, *40*(3), 625–645.

Jung, J., Jayakrishnan, R., & Park, J. Y. (2013). Design and modeling of real-time shared-taxi dispatch algorithms. Presented at the 92nd Annual MeetingTransportation Research Board. Retrieved from https://academic.microsoft.com/#/detail/305515526

Kamga, C., Yazici, M. A., & Singhal, A. (2013). Hailing in the rain: Temporal and weather-related variations in taxi ridership and taxi demand-supply equilibrium. In *Transportation Research Board 92nd Annual Meeting*.

Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical Methods versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights. *Transportation Research Part C: Emerging Technologies*, *19*(3), 387–399.

Karner, A. (2018). Assessing public transit service equity using route-level accessibility measures and public data. *Journal of Transport Geography*, *67*, 24–32.

Kattan, L., de Barros, A., & Wirasinghe, S. C. (2010). Analysis of work trips made by taxi in canadian cities. *Journal of Advanced Transportation*, *44*(1), 11–18.

Keijer, M. J. N., & Rietveld, P. (2000). How do people get to the railway station? The dutch experience. *Transportation Planning and Technology*, *23*(3), 215–235. https://doi.org/10.1080/03081060008717650

Kim, S., & Ulfarsson, G. (2004). Travel mode choice of the elderly: effects of personal, household, neighborhood, and trip characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, (1894), 117–126.

King, D. A., Peters, J. R., & Daus, M. W. (2012). Taxicabs for Improved Urban Mobility: Are We Missing an Opportunity? In *Transportation Research Board 91st Annual Meeting*. Retrieved from http://trid.trb.org/view.aspx?id=1129618

King, D. A., & Saldarriaga, J. F. (2016). Access to Taxicabs for Unbanked Households: An Exploratory Analysis in New York City. In *Transportation Research Board 95th Annual Meeting*. Retrieved from https://trid.trb.org/view.aspx?id=1394186

Kitamura, R., Mokhtarian, P. L., & Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, *24*(2), 125–158. https://doi.org/10.1023/A:1017959825565

Klöckner, C. A., & Friedrichsmeier, T. (2011). A multi-level approach to travel mode choice – How person characteristics and situation specific aspects determine car use in a student sample. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(4), 261–277. https://doi.org/10.1016/j.trf.2011.01.006

Knight, T. E. (1974). An approach to the evaluation of changes in travel unreliability: a "safety margin" hypothesis. *Transportation*, *3*(4), 393–408.

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board*, (1607), 116–125.

Kunkle, F. (2018, May 26). Uber, Lyft and other ride-hailing apps are 'useless' for disabled riders, NYC advocates say. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/tripping/wp/2018/05/26/uber-lyft-and-other-ride-hailing-apps-are-useless-for-disabled-riders-nyc-advocates-say/

Kuppam, A. R., & Pendyala, R. M. (2001). A structural equations analysis of commuters' activity and travel patterns. *Transportation*, *28*(1), 33–54.

Lacombe, A., & Morency, C. (2016). Modeling Taxi Trip Generation Using GPS Data: The Montreal Case. In *Transportation Research Board 95th Annual Meeting*.

Lambert, O. (2016a, February 1). A new report explores the benefits of Uber. Retrieved from http://www.news.com.au/finance/business/travel/a-new-report-explores-the-benefits-of-uber/news-story/5e41687c9bc2614f07015f85eb784f7f

Lambert, O. (2016b, February 1). Is Uber better than taxis? Retrieved October 7, 2016, from http://www.news.com.au/finance/business/travel/a-new-report-explores-the-benefits-of-uber/news-story/5e41687c9bc2614f07015f85eb784f7f

Li, Y., Xia, T., & Duan, H. (2014). The Impact on Taxi Industry of Taxi-Calling Mobile Apps in Shanghai. In *Transportation Research Board 93rd Annual Meeting*. Retrieved from http://trid.trb.org/view.aspx?id=1289065

Limtanakool, N., Dijst, M., & Schwanen, T. (2006). The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium-and longer-distance trips. *Journal of Transport Geography*, *14*(5), 327–341.

Litman, T. (2003). Measuring transportation: Traffic, mobility and accessibility. *Institute of Transportation Engineers. ITE Journal*, *73*(10), 28–32.

Lund, H. M., Cervero, R., & Wilson, R. W. (2004). *Travel characteristics of transit-oriented development in California*. publisher not identified. Retrieved from http://staging.community-wealth.org/sites/clone.community-wealth.org/files/downloads/report-lund-cerv-wil.pdf

Mahesh, R. (2015). From Jitney Buses to App-Based Ridesharing: Understanding California's' Third Way'Approach to Ride-for-Hire Regulation. *Southern California Law Review*, *88*(4). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474452

Matthies, E., Kuhn, S., & Klöckner, C. A. (2002). Travel Mode Choice of Women The Result of Limitation, Ecological Norm, or Weak Habit? *Environment and Behavior*, *34*(2), 163–177. https://doi.org/10.1177/0013916502034002001

McDonald, N. C. (2008). Critical Factors for Active Transportation to School Among Low-Income and Minority Students: Evidence from the 2001 National Household Travel Survey. *American Journal of Preventive Medicine*, *34*(4), 341–344. https://doi.org/10.1016/j.amepre.2008.01.004

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*(4), 303–328. https://doi.org/10.1016/0047-2727(74)90003-6

Meadows, A. (2015, June 26). Is Uber cheaper than a taxi? Not always. *Christian Science Monitor*. Retrieved from http://www.csmonitor.com/Business/Saving-Money/2015/0626/Is-Uber-cheaper-than-a-taxi-Not-always

Metcalfe, G., & Warburg, J. (2012). A Policy Agenda for the Sharing Economy. Retrieved February 3, 2016, from http://www.spur.org/publications/article/2012-10-09/policy-agenda-sharing-economy

Miller, E. J., Roorda, M. J., & Carrasco, J. A. (2005). A tour-based model of travel mode choice. *Transportation*, *32*(4), 399–422. https://doi.org/10.1007/s11116-004-7962-3

Mishalani, R., McCord, M., & Wirtz, J. (2006). Passenger Wait Time Perceptions at Bus Stops: Empirical Results and Impact on Evaluating Real - Time Bus Arrival Information. *Journal of Public Transportation*, *9*(2). http://dx.doi.org/10.5038/2375-0901.9.2.5

Mokhtarian, P. L. (2016). Presenting the Independence of Irrelevant Alternatives property in a first course on logit modeling. *Journal of Choice Modelling*, *21*, 25–29.

Mokhtarian, P. L., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological*, *42*(3), 204–228.

Morang, M. Z. (2014). Add GTFS to a Network Dataset | GTFS in ArcGIS. Retrieved June 2, 2018, from https://esri.github.io/public-transit-tools/AddGTFStoaNetworkDataset.html

Murakami, E., & Young, J. (1997). *Daily travel by persons with low income*. US Federal Highway Administration Washington, DC. Retrieved from http://nhts.ornl.gov/1995/Doc/lowinc.pdf

Nassir, N., Hickman, M., Malekzadeh, A., & Irannezhad, E. (2016). A utility-based travel impedance measure for public transit network accessibility. *Transportation Research Part A: Policy and Practice*, *88*, 26–39.

Newman, J. P., & Bernardin, V. L. (2010). Hierarchical ordering of nests in a joint mode and destination choice model. *Transportation*, *37*(4), 677–688.

Noland, R. B., & Polak, J. W. (2002). Travel time variability: a review of theoretical and empirical issues. *Transport Reviews*, *22*(1), 39–54.

Noland, R. B., & Small, K. A. (1995). *Travel-time uncertainty, departure time choice, and the cost of the morning commute*. Institute of Transportation Studies, University of California, Irvine. Retrieved from http://128.200.36.2/its/publications/papers/ITS/UCI-ITS-WP-95-1.pdf

Nurul Habib, K. M. (2012). Modeling commuting mode choice jointly with work start time and work duration. *Transportation Research Part A: Policy and Practice*, *46*(1), 33–47. https://doi.org/10.1016/j.tra.2011.09.012

Nutley, S. (2005). Monitoring rural travel behaviour: a longitudinal study in Northern Ireland 1979–2001. *Journal of Transport Geography*, *13*(3), 247–263.

Omrani, H., Charif, O., Gerber, P., Awasthi, A., & Trigano, P. (2013). Prediction of Individual Travel Mode with Evidential Neural Network Model. *Transportation Research Record: Journal of the Transportation Research Board*, *2399*, 1–8. https://doi.org/10.3141/2399-01

Parfenov, S., Weeks, A., & Alam, Z. (2012). Travel Patterns of NYC Yellow Taxis: Routing, Activity, and Results. Retrieved from http://proceedings.esri.com/library/userconf/proc14/papers/388_540.pdf

Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A., & Ratti, C. (2010). Taxi-aware map: Identifying and predicting vacant taxis in the city. In *Ambient Intelligence* (pp. 86–95). Springer.

Picci, A. (2016). Uber vs. Taxi: Which Is Cheaper? Retrieved October 8, 2016, from http://www.consumerreports.org/personal-finance/uber-vs-taxi-which-is-cheaper/

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., & Waddell, P. A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, *38*(6), 933.

Plaut, P. O. (2005). Non-motorized commuting in the US. *Transportation Research Part D: Transport and Environment*, *10*(5), 347–356.

Polzin, S., Chu, X., & Rey, J. (2000). Density and captivity in public transit success: observations from the 1995 nationwide personal transportation study. *Transportation Research Record: Journal of the Transportation Research Board*, (1735), 10–18.

Poon, L. (2015). A New NYC Taxi App Shows There's Still Hope for Cabs. Retrieved June 26, 2018, from http://www.citylab.com/commute/2015/08/a-new-nyc-taxi-app-shows-theres-still-hope-for-cabs/402577/

Pucher, J., & Buehler, R. (2009). Integrating bicycling and public transport in North America. *Journal of Public Transportation*, *12*(3), 5.

Pucher, J., & Renne, J. L. (2003). Socioeconomics of urban travel: evidence from the 2001 NHTS. *Transportation Quarterly*, *57*(3), 49–77.

Qian, X., & Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. *Applied Geography*, *59*, 31–42.

Qian, X., Zhan, X., & Ukkusuri, S. V. (2015). Characterizing Urban Dynamics Using Large Scale Taxicab Data. In *Engineering and Applied Sciences Optimization* (pp. 17–32). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-18320-6_2

Rajamani, J., Bhat, C., Handy, S., Knaap, G., & Song, Y. (2003). Assessing impact of urban form measures on nonwork trip mode choice after controlling for demographic and level-of-service effects. *Transportation Research Record: Journal of the Transportation Research Board*, (1831), 158–165.

Rao, P. S., Sikdar, P. K., Rao, K. K., & Dhingra, S. L. (1998). Another insight into artificial neural networks through behavioural analysis of access mode choice. *Computers, Environment and Urban Systems*, *22*(5), 485–496.

Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, *45*, 168–178. https://doi.org/10.1016/j.tranpol.2015.10.004

Reilly, M., & Landis, J. (2003). The influence of built-form and land use on mode choice. *University of California Transportation Center*. Retrieved from https://escholarship.org/uc/item/46r3k871.pdf

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D: Transport and Environment*, *5*(1), 71–75. https://doi.org/10.1016/S1361-9209(99)00019-X

Rodríguez, D. A., & Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D: Transport and Environment*, *9*(2), 151–173.

Rosenbloom, S. (2003). Increasing accessible taxi options for people with disabilities. In *82 nd TRB Annual Meeting, Washington, DC*.

Rosenbloom, Sandra. (2004). Mobility of the Elderly. *Transportation in an Aging Soceity*, 3–21.

Salnikov, V., Lambiotte, R., Noulas, A., & Mascolo, C. (2015). OpenStreetCab: Exploiting Taxi Mobility Patterns in New York City to Reduce Commuter Costs. *ArXiv Preprint ArXiv:1503.03021*. Retrieved from http://arxiv.org/abs/1503.03021

Saneinejad, S., Roorda, M. J., & Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport and Environment*, *17*(2), 129–137.

Schaller, B. (2005). A regression model of the number of taxicabs in US cities. *Journal of Public Transportation*, *8*(5), 4.

Scheiner, J., & Holz-Rau, C. (2007). Travel mode choice: affected by objective or subjective determinants? *Transportation*, *34*(4), 487–511.

Schmöcker, J. D., Quddus, M. A., Noland, R. B., & Bell, M. G. (2008). Mode choice of older and disabled people: a case study of shopping trips in London. *Journal of Transport Geography*, *16*(4), 257–267.

Schmöcker, J. D., Quddus, M. A., Noland, R. B., & Bell, M. G. H. (2007). Transport mode choice of the elderly and disabled in London. *Transportation Geography*.

Schmöcker, J. D., Quddus, M., Noland, R., & Bell, M. (2005). Estimating Trip Generation of Elderly and Disabled People: Analysis of London Data. *Transportation Research Record: Journal of the Transportation Research Board*, (1924), 9–18.

Schwanen, T., & Mokhtarian, P. L. (2005). What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*, *13*(1), 83–99.

Sekhar, C. R., Minal, & Madhu, E. (2016). Mode Choice Analysis Using Random Forrest Decision Trees. *Transportation Research Procedia*, *17*, 644–652. https://doi.org/10.1016/j.trpro.2016.11.119

Shaheen, S. A., Chan, N. D., & Micheaux, H. (2015). One-way carsharing's evolution and operator perspectives from the Americas. *Transportation*, *42*(3), 519–536.

Shiftan, Y. (1998). Practical approach to model trip chaining. *Transportation Research Record: Journal of the Transportation Research Board*, (1645), 17–23.

Shukla, N., Ma, J., Wickramasuriya, R., Huynh, N. N., & Perez, P. (2015). Tour-based travel mode choice estimation based on data mining and fuzzy techniques. Retrieved from http://ro.uow.edu.au/cgi/viewcontent.cgi?article=5345&context=eispapers

Small, K. A. (1982). The scheduling of consumer activities: work trips. *The American Economic Review*, *72*(3), 467–479.

Smart, R., Rowe, B., Hawken, A., Kleiman, M., Mladenovic, N., Gehred, P., & Manning, C. (2015). Faster and Cheaper: How Ride-Sourcing Fills a Gap in Low-Income Los Angeles Neighborhoods. *BOTEC Analysis Corporation. Retrieved September*, *17*, 2015.

Song, L., Zhang, D., Chen, J., An, J., & Su, Y. (2008). Analysis of taxi operation characteristics with traffic control. *Journal of Transportation Systems Engineering and Information Technology*, *8*(6), 127–131.

Sorrel, C. (2015, November 24). Ride-Hailing Services Like Uber Are Terrible For Disabled People. Retrieved June 26, 2018, from https://www.fastcompany.com/3052721/ride-hailing-services-like-uber-are-terrible-for-disabled-people

Stopher, P. (1969). A Multinomial Extension of the Binary Logit Model for Choice of Mode of Travel. *Northwestern University, Unpublished*.

Su, F., & Bell, M. G. (2009). Transport for older people: Characteristics and solutions. *Research in Transportation Economics*, *25*(1), 46–55.

Taylor, D., & Mahmassani, H. (1996). Analysis of stated preferences for intermodal bicycle-transit interfaces. *Transportation Research Record: Journal of the Transportation Research Board*, (1556), 86–95.

Verto Analytics. (2018). Monthly number of Uber and Lyft users in the United States from January to March 2018 (in millions). Retrieved June 21, 2018, from https://www.statista.com/statistics/833743/us-users-ride-sharing-services/

Vythoulkas, P. C., & Koutsopoulos, H. N. (2003). Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. *Transportation Research Part C: Emerging Technologies*, *11*(1), 51–73.

Wang, F., & Ross, C. L. (2017). New potential for multimodal connection: exploring the relationship between taxi and transit in New York City (NYC). *Transportation*, 1–22.

Wear, B. (2018). Why some say new Austin taxi proposal would be bad for business. Retrieved June 26, 2018, from https://www.mystatesman.com/news/local/why-some-say-new-austin-taxi-proposal-would-bad-for-business/2XavaPGT8M1t9mxMqylbLL/

Welch, T. F., Gehrke, S. R., & Wang, F. (2016). Long-term impact of network access to bike facilities and public transit stations on housing sales prices in Portland, Oregon. *Journal of Transport Geography*, *54*, 264–272.

Wets, G., Vanhoof, K., Arentze, T., & Timmermans, H. (2000). Identifying decision structures underlying activity patterns: an exploration of data mining algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, (1718), 1–9.

Xie, C., Lu, J., & Parkany, E. (2003). Work travel mode choice modeling with data mining: decision trees and neural networks. *Transportation Research Record: Journal of the Transportation Research Board*, (1854), 50–61.

Yang, C., & Gonzales, E. (2014). Modeling Taxi Trip Demand by Time of Day in New York City. *Transportation Research Record: Journal of the Transportation Research Board*, (2429), 110–120.

Ye, X., Pendyala, R. M., & Gottardi, G. (2007). An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B: Methodological*, *41*(1), 96–113.

Zhang, M. (2004). The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association*, *70*(3), 344–360. https://doi.org/10.1080/01944360408976383

Zhang, W., Guhathakurta, S., Fang, J., & Zhang, G. (2015). Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. *Sustainable Cities and Society*, *19*, 34–45. https://doi.org/10.1016/j.scs.2015.07.006

Zhang, Y., & Xie, Y. (2008). Travel mode choice modeling with support vector machines. *Transportation Research Record: Journal of the Transportation Research Board*, (2076), 141–150.

Zhen, C. (2015). *Impact of Ride-Sourcing Services on Travel Habits and Transportation Planning*. University of Pittsburgh. Retrieved from http://d-scholarship.pitt.edu/25827/

Zielinski, S. (2016). MaaS APPEAL: the Meta Transformation of Mobility Arrives in Michigan. Retrieved October 7, 2016, from https://www.nextenergy.org/maas-appeal-meta-transformation-mobility-arrives-michigan/